Ability Grouping in the Public Schools. A Field Study
Author(s): Walter R. Borg
Source: *The Journal of Experimental Education*, Vol. 34, No. 2, Ability Grouping in the Public Schools. A Field Study (Winter, 1965), pp. 1-97
Published by: Taylor & Francis, Ltd.
Stable URL: http://www.jstor.org/stable/20156816
Accessed: 12-07-2016 22:32 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Taylor & Francis, Ltd. is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Experimental Education*

CHAPTER 1

INTRODUCTION: THE FIELD STUDY

THE GENERAL GOAL of this research project has been to study differences in the effects upon elementary, junior high school, and high school pupils of an ability grouping system that distinguished between a curriculum that adjusted the rate of presentation of curricular materials, and a random grouping system that differentiated the curriculum principally through the use of enrichment. The study explores a wide range of dependent variables that might be affected differentially by the two grouping treatments, including many not previously explored in ability grouping research.

Scope of the Project

The paper has been divided into three sections. The first section contains a brief description of the research procedures including the samples studied, the measures used, and the ability grouping and random grouping treatments used by the two participating districts. The second section is concerned with the achievement-related dependent variables in the two grouping treatments including general achievement, over and underachievers and study methods. The third section discusses what we might call non-cognitive dependent variables that could be influenced by the grouping treatments such as the pupil's social adjustment, attitudes toward school, problems, self-concept, and personality.

Perhaps the most important variable to be considered in studies involving instructional methods is achievement. This research has explored the question of achievement differences in ability grouped and random grouped classrooms more thoroughly than any work previously published, considering a number of achievement areas over a four-year period, and measuring achievement differences at all grade levels from fourth through twelfth.

A careful study was also made to determine whether study habits and methods differed under the two grouping treatments. This research explored, not only the claim often made by educators that ability grouping leads to better study methods among bright pupils, but also analyzed differences and changes in study habits for average and slow pupils in the two systems.

The possible effects of ability grouping upon social adjustment and pupil peer status has also been a cause for concern among educators. This research has attempted to answer some of the questions in this area by studying differences in sociometric status between comparable pupils in ability grouped and random grouped classrooms.

Some educators have condemned ability grouping on the grounds that it is destructive to the individual's self concept. It has been suggested that in the ability grouped classroom bright pupils become arrogant, while slow pupils become insecure and defeated. The slow pupil's low opinion of himself, it has been suggested, leads him to adjust his goals and efforts downward to fit the label place on him in the ability grouping situation. This research examines in considerable detail these and other questions relative to self-concept differences in ability grouped and random grouped classrooms.

Somewhat related to self-concept are a number of personality areas that may also be affected by the ability grouping situation. Personality measures, both of the group administered inventory type and the individually administered projective type, were employed in this research to explore the possible effects of ability grouping upon certain personality variables.

The attitudes of pupils toward the school environment and their adjustments to it were also explored in this research. This area has received little attention in previous research but appears to be an important consideration in comparing the ability grouping and random grouping treatments.

Research Procedures

Two adjacent and closely comparable school districts in Utah provided the setting in which the above questions were explored. One of these districts employed random grouping with enrichment, and the other had adopted a system of ability grouping with acceleration, coincident with the start of this research. Over 2,500 pupils from the fourth, sixth, seventh, eighth, and ninth grade levels were selected in the two districts at the beginning of the study. During the second year of the study, this sample was increased to about 4,000 pupils. Research data were collected over a four year period in order to appraise the long term effects of the two grouping treatments. Pupils who were first tested early in grade four were followed through grade seven. As all samples were similarly followed over the period of the study, data were collected at all grades from four through twelve.

The district that employed ability grouping, referred to throughout this report as District A, used a composite achievement test score as a basis for grouping elementary school pupils into self-contained classrooms at the elementary level. At the secondary level, ability grouping was employed in mathematics and science classes only. Pupils were grouped into three ability levels referred to in this report as superior, average, and slow, in both the elementary and secondary schools.

The other district participating in this research, known as District R, employed random or heterogeneous grouping and differentiated the curriculum for individual differences in the random grouped classroom through the use of enrichment. The District A and District R curricula were very similar in most subject areas and at most grade levels, and in many instances the materials employed in District A to adjust the rate of presentation of the curriculum were also used in District R to provide enrichment in heterogeneous classes. This project called for comparisons between pupils in the two grouping treatments who were of comparable ability level. Pupils in District A who had been assigned to the three types of ability groups were measured by the investigators and then, after administering the same measure to District R pupils, the latter were classified into three ability levels so as to provide samples of pupils of comparable level in the two districts.

Except during the pilot study year, the measures employed in this study were not part of the regular testing program of either district and were administered by teams of examiners trained and supervised by the principal investigator. These examiners were employed by Utah State University and were not affiliated with either school district.

Analysis

Data collected in this research were analyzed using several procedures. In cases where it was desirable to measure changes that had taken place in pupils who had been exposed to the two grouping treatments, analysis of covariance was used. For example, this technique was employed in studying the achievement data. In this case, the final achievement scores of the various groups studied were adjusted to compensate for differences present at the beginning of the period covered by the analysis. Analysis of variance was employed to compare subjects on most of the other dependent variables. Variables that were analyzed using this technique included those for which it appeared reasonable to assume that the sampling techniques would lead to initially comparable groups, and those for which an initial measure could not be obtained.

For other portions of the research data, chi-square analysis was employed. This technique was generally used in cases where the data fell into a number of well-defined catagories or in cases where parametric statistics did not appear appropriate. Much of the sociometric data were analyzed using chi-square contingency tables because this permitted use of the standard sociometric classifications (stars, isolates, etc.) that are generally understood by persons familiar with this area of measurement.

Three major types of differences were considered in analyzing most of the data. The first and most important comparisons were between comparable pupils in the two grouping treatments. These comparisons involved pupils of the same sex and ability level in the ability grouped and random grouped samples and are referred to throughout the study as between-treatment differences. These differences provide the best evidence on the differential effects of ability grouping and random grouping upon the dependent variables.

Comparisons between pupils of the same sex and in the same grouping treatment who differed in ability level are referred to as level or within-treatment differences. These comparisons tell us whether the grouping treatments had a differential effect upon pupils at different ability levels and give us additional insight about the treatment that help to interpret the between-treatment differences. Boys and girls who were in the same grouping treatment and at the same ability level were compared to provide knowledge about the sex-related differences on the dependent variables. In some instances, these sex differences also provided additional insights into the effects of the grouping treatment.[2]

THE SAMPLE

As is the case with most field studies in education, random samples of broadly defined populations could not be obtained in the research herein described. Even if such a random sample had been available, the problem of randomly dividing such a sample and exposing half of them to a school situation involving ability grouping and the other half to random grouping would have been nearly impossible. Random samples of more narrowly defined populations such as "all pupils enrolled in regular fourth grade classes in school district X" can be obtained in field studies, but even this approach often causes administrative problems that public school officials consider too great a disruption of the school's activities. Thus, the educational field study must usually make compromises in sample selection that would be unacceptable in the more rigorous setting of the psychological laboratory.

Yet, the field study has certain advantages that make it a useful approach in educational research. First, although the discipline of the laboratory is sacrificed, the artificiality that usually attends it is greatly reduced in the field study. Thus, the field study gathers data in realistic and reasonably typical school situations and its findings are often directly applicable to other school situations. In contrast, the laboratory school which employs highly selected teachers to work with small classes unhampered by financial limitations and free to draw professional support from a university faculty, often produces research results that have little carryover to the usual school in which classes are large, teachers are average, and money is scarce. Second, the field study can work with larger numbers of pupils than are usually manageable in laboratory studies. This permits identifying and studying subgroups that would be too small to consider with the usual laboratory sample. More important, these large samples offer a greater possibility than variables one cannot control, and, for that matter, variables that we are not even aware of will "randomize out" or tend to occur equally in the groups we wish to compare.

The researcher doing a field study, though usually unable to work with truly random samples, is still obligated to make his samples as random as the situation permits and to be alert to biasing elements in his sampling procedure that may distort his findings. His second obligation is to describe with reasonable thoroughness the situation in which the research was carried out. Unless this is done, the reader has little upon which to base his evaluation of the project and his estimate of the pertinence of the findings to other pupils in other schools. The nature of the schools involved, the comparability of their curricula, the numbers of pupils in the sample and how they were selected are among the topics to be discussed in this section.

District Comparisons

District R, from which the random grouped sample was selected, is a city school district having an average daily attendence of about 17,000 at the time of the study. District A, from which the ability grouped sample was selected, is the surrounding county school district with an ADA of about 12,000. Similarities and differences between these districts must be considered in interpreting the results of this research. It is impossible, of course, to compare the two districts on all aspects of the educational environment they provide and the population they serve. Certain critical comparisons, however, have been made and are reported throughout this section.

In terms of financial resources for instruction, the two districts were closely comparable during the four years of this study. The yearly average amount available for instruction was $238.76 per pupil in District A and $230.48 in District R. When it is considered that District A found it necessary to purchase a number of materials in order to fit their curriculum to the ability grouping system, it would seem appropriate to conclude that financial resources available for instruction were closely comparable.

Perhaps the most important element to consider is the relative skill of teachers in the ability grouped and random grouped classrooms from which the research samples were selected. Unfortunately, a valid measurement of teaching ability is extremely difficult if not impossible at our present state of knowledge. Thus, our comparisons between teachers must rely on variables which, though indirect, are probably valid indicators of the skill of a given group of teachers.

Salaries paid to participating teachers in the two districts were very similar with the first year mean for District A being $4744 as compared with a District R mean of $4788. Teachers in the two districts were also closely comparable in terms of certification, but in terms of teaching experience a significant difference existed with District R teachers having on the average about four and a half years more experience than District A teachers. This difference existed at both secondary and elementary school levels. All in all, however, teachers in the two districts were reasonably comparable with neither district showing a consistent superiority on the variables compared.

Another important factor that must be considered in comparing Districts A and R is the curriculum. Major differences in the basic curricula of the two districts would make achievement comparisons meaningless because any achievement measure used would then be more appropriate for one district than the other. A study in which the same curriculum was adjusted by the two different treatments – ability grouping with acceleration and random grouping with enrichment – would provide an ideal situation for comparing the treatments. The basic curricula in Districts A and R, though not identical, were very similar and provided a sound basis for comparing the two teaching approaches with which this research is concerned. In the State of Utah, most school subjects are covered by curriculum guides that have been developed cooperatively by the Utah State Department of Public Instruction and the school districts of the state. These guides are published by the state and were followed by both of the participating school districts; not only in the subject areas that were measured in this research, but also in most other subjects taught in the elementary and secondary schools. Both districts also selected their textbooks from the state approved textbook list which is made up of comparable textbooks meeting the objectives of the state curriculum guides. In some subject areas, the two districts used the same textbooks. Thus, the curricula offered by the two districts were very similar in terms of major objectives and sequences. In many cases, the basic District A curriculum which was geared to pupils in the average ability group, was found to be almost identical in rate, sequence, and content to the District R basic curriculum for heterogeneous groups. Differences did occur, however, between the District A and R curricula for superior and slow pupils, largely because of the considerable adjustments in rate of instruction made in District A. For example, nearly all District A superior pupils were scheduled to take Algebra 1 in the eighth grade; while District R pupils did not take Algebra 1 until the ninth grade. As enrichment materials were available in District R classrooms, one would expect that the District R superior pupils, while moving at a slower pace than their counterparts in District A, would have more time for independent study and projects and might develop greater depth in the subjects covered.

Selecting Research Subjects

During the pilot study year of the project, subjects were selected from the fourth, sixth, seventh, eighth, and ninth grades. Table 1 gives the number of pupils in each of these five samples during the four years of the project. These data are based on numbers of pupils completing each achievement test battery.

Because a fairly large number of variables were explored in this research, the numbers of pupils for whom complete data are available changes from one variable to another. For example, a pupil may have completed all tests relating to personality or self-concept but may have missed and been unable to make up one of the achievement tests. He may, therefore, be included in the self-concept and personality analysis but not in parts of the achievement analysis. The number of students available for the achievement analysis was generally smallest because the achievement batteries were longer, were usually administered over a period of four days at the elementary level and two days at the secondary level and were thus more difficult to make up. In general, however, this table gives the reader a reasonable picture of the numbers of pupils involved in the Utah study.

Initial samples were drawn in classroom units rather than by individuals. This procedure, of course, is less satisfactory than random selection of individual cases.

Pupils who were in the fourth grade at the start of the project (1958-59), are referred to as Sample IV throughout this report. This fourth-grade sample in District A came from nine classrooms in three elementary schools. Three superior sections, three average sections and three slow sections were included, one of each from each school. The District R sample came from 13 classrooms in six schools. These schools were selected for comparability with the District A schools in terms of socioeconomic status of the neighborhood served. Classrooms were selected at random from among the available classrooms in each school. As District R pupils were randomly grouped, a larger sample was considered necessary in order to insure getting enough subjects at each ability level. Prior to the beginning of the second year of the study, support was obtained from the Cooperative Research Program, U. S. Office of Education, to carry out the remaining three years of the project. At this time the number of pupils in the fourth grade samples was increased.

Pupils selected from the sixth grade during the initial year of the study have been designated as Sample VI. The District A Sample VI was taken from 12 classrooms in four schools and included four superior sections, four average sections, and four slow sections. The sixth grade sample in District R was taken from 16 classrooms in eight schools. The procedure for selecting sixth grade classrooms was the same as for fourth grade classrooms.

Seventh grade pupils selected during the initial year of the study are designated as Sample VII. During the pilot study year, the District A sample consisted of all pupils in three seventh grade mathematics classes, one each being randomly selected from the superior, average and slow classes. The Sample VII subjects for District R during the first year were taken from seven randomly selected seventh grade mathematics classes. These samples were substantially enlarged during the second year of the study.

Eighth grade pupils selected during the initial year are designated as Sample VIII. These pupils, as was the case with Sample VII, were taken from the junior high schools that ordinarily serve the schools from which the elementary samples were drawn. District A Sample VIII consisted of all pupils in six eighth grade mathematics classes; two randomly selected from each ability level. District R Sample VIII consisted of all pupils in seven randonly selected eighth grade mathematics classes. Both samples were enlarged during the second year of the study.

Pupils in the ninth grade during the initial year of the study are designated as Sample IX. These pupils were studied from the beginning of the ninth grade through the end of the twelfth grade. In District A, this sample included all pupils in three ninth grade algebra classes, and three ninth grade science classes, one in each subject at slow, average, and superior levels. Because many pupils at the superior and slow levels were in both the mathematics and science classes selected, this sample was quite small. In District R, Sample IX included all pupils taken from four randomly selected ninth grade mathematics classes.

TABLE 1

NUMBER OF PUPILS IN EACH SAMPLE[a]

	Fall 1958 (CAT)		Spring 1960 (STEP)		Spring 1961 (STEP)[b]		Spring 1962 (STEP)	
	District A	District R	District A	District R	District A	District R	District A	District R
Sample IV								
Superior	100	145	134	253	85	228	82	231
Average	110	151	228	220	141	159	138	159
Slow	83	78	86	110	43	81	43	82
Total	293	374	448	583	269	468	263	472
Sample VI								
Superior	172	215	168	217	157	195	154	184
Average	143	140	284	134	258	117	243	115
Slow	126	79	126	77	83	82	104	70
Total	441	434	578	428	498	394	501	369
Sample VII								
Superior	24	72	148	143	110	125	100	120
Average	47	103	259	129	235	104	214	86
Slow	20	38	51	61	55	54	41	42
Total	91	213	458	333	400	283	355	248
Sample VIII								
Superior	69	80	110	151	104	121	97	115
Average	78	57	216	71	185	54	161	55
Slow	32	61	54	72	39	51	37	42
Total	179	198	380	294	328	226	295	212
Sample IX								
Superior	33	92	33	92	27	75	23	65
Average	88	38	85	38	73	30	69	26
Slow	15	41	14	41	13	30	12	24
Total	136	171	132	171	113	135	104	115
Yearly Total	1140	1390	1996	1809	1608	1506	1518	1416

[a] Based on achievement test batteries as indicated; *California Achievement Tests* (CAT) or *Sequential Tests of Educational Progress* (STEP)

[b] No achievement battery was administered at secondary level this year; size of Sample VI, VII, VIII and IX based on the number completing the *Index of Adjustment and Values*.

Comparability of Research Subjects

Although considerable effort was made to avoid techniques that would lead to biasing of the samples, the techniques used were far from ideal. Thus, during the course of the study, and especially in the pilot study year, data were collected to check our success in selecting reasonably comparable subjects. Some of these data will now be presented.

Socioeconomic Class: It will be recalled that elementary schools were selected at the outset of the study that were considered to serve comparable socioeconomic groups in the two districts. A preliminary check of socioeconomic status made in these schools indicated no significant differences. It was considered desirable, however, to make a more thorough check of socioeconomic differences as this variable could have a considerable effect upon the outcome of the research. The Hollingshead Occupational Scale was selected as a basis for establishing five socioeconomic class levels. This scale is based entirely upon the occupation of the father or the principal family wage earner. Although much easier to obtain than some of the more elaborate socioeconomic class measures such as Warner's Index of Status Characteristics, the Hollingshead Occupational Scale has been found to correlate highly with such measures (61). These data were collected during the final year of the study. At this time, Sample IV and Sample VI pupils were in the seventh and ninth grades respectively. At this level they were able to provide more information on their father's occupations than they could have at the elementary school level. The distribution of cases into the five Hollingshead socioeconomic class levels was compared using chi square contingency tables to determine whether pupils of comparable ability levels in the two districts were significantly different in their distribution into the five social class levels. Of the six chi square values calculated for Samples IV and VI, one was statistically significant. The significant chi square was found between District A and District R superior pupils in the Sample VI. This chi square reached significance at the. 01 level and indicated a tendency for District R superior pupils to be of higher socioeconomic status. None of the other chi square values reached significance at the . 05 level and most were small, indicating no true differences in socioeconomic distribution between the two samples. The significant difference, however, between superior Sample VI pupils in Districts A and R should be taken into consideration in interpreting differences found between these groups on the other variables considered in this research.

Intelligence Differences: Comparisons of mean intelligence levels for District A and District R show that most of the parallel groups in the two districts were closely comparable in mean mental age expressed in grade placement equivalents. For the language score on the California Short-Form Test of Mental Maturity, the average difference between comparable ability groups in Districts A and R was less than two-tenths of a grade.

The differences between comparable District A and District R pupils were somewhat larger on the CMM non-language score, averaging nearly three-tenths of a grade. The greatest difference was found between the slow pupils in Sample IX and favored the District R pupils. Even this largest difference represents less than two raw score points (28. 95 vs. 27. 25) and falls far short of statistical significance.

THE TREATMENT

Because terms such as "ability grouping", "enrichment" and "heterogeneous grouping" have different meaning in different situations, it is necessary to define operationally their meaning in a research project so that the reader fully understands the specific circumstances under which the data were collected.

The Ability Grouping System

The ability grouping system used in District A during the course of this research employed a standard achievement battery, the California Achievement Test Battery (CAT), as a basis for grouping pupils into three ability levels. An achievement measure was selected instead of scholastic aptitude or intelligence measures largely because the district administration found that grouping on the basis of past achievement was more acceptable to parents than grouping on the basis of ability. Scores on a battery of achievement tests have generally been found to correlate rather highly with measures of scholastic aptitude, so it is doubtful if any marked difference in the grouping would grow out of a change from one of these measures to another.

At the elementary school level, the total grade placement score on the California Achievement Test Battery, Form W, was used as a basis for grouping. Pupils were administered the test battery in the spring of each school year. Pupils at a given grade level were then divided into boys and girls and a separate list made up for each sex starting with the pupil having the highest grade placement score and proceeding in order to the pupil with the lowest score. The lowest ten boys and lowest ten girls were then selected and were placed in the slow section. Next, the highest 15 boys and highest 15 girls were selected and were placed in the superior section. The remaining pupils were placed in two or three average sections by placing every second or third pupil in a given section. Thus, the average sections within a given school were comparable in terms of ability level and range rather than being divided into "high average" and "low average" sections. Both districts had a separate program of special classes for mentally retarded pupils, so these were not included in either the ability grouped or random grouped classes. A number of special provisions were made for unusual situations such as an excessive number of superior or slow pupils in a given school, too large an ability gap between boys and girls assigned to a slow section, and others that seemed to call for special treatment. This simple three-level grouping system based on a composite achievement score has both advantages and disadvantages. Among its advantages are that it may be carried out in relatively small schools, it is easy to use and generally acceptable to parents, and it permits a relatively easy transition for pupils transferring to or from schools employing random grouping. Finally, this system permits the use of self-contained classrooms at the elementary level. When specific grouping criteria are established on a subject-by-subject basis, some sort of departmentalized program is usually called for, as pupils will sometimes fall into different ability groups for different subject areas.

The principal disadvantage of the grouping system employed in District A at the elementary level is that the use of a general grouping criterion such as a composite achievement score results in only a moderate reduction in heterogeneity. It does not produce pupil groups that are at or even near the same level on all subjects included in the curriculum. Thus District A employed what might be referred to as a conservative grouping system that can be adapted to most elementary schools, but brings about only moderate reductions in heterogeneity.

At the secondary school level, District A employed ability grouping in mathematics and science classes only. Even within these two subject areas, the classes that could be satisfactorily grouped became fewer as the pupil moved into the higher grade levels. The grouping criterion for mathematics was the grade placement score of the Mathematics subtest of the California Achievement Test Battery. The achievement score obtained in the spring on this subtest formed the basis for grouping pupils into different ability levels the following fall. In science, on the other hand, since the CAT has no science subtest, the composite grade placement score for the Mathematics and Language subtests was used, thus providing a general criterion for grouping in this subject.

Ability grouping was employed in all mathematics and science classes during the junior high-school grades. At the high school level, ability grouping was only possible in the lower level mathematics and science courses that are taken by a sizable percentage of pupils. In such elective courses as physics and chemistry, the maximum amount of grouping possible was at two levels since virtually no slow pupils took these elective physical science courses.

The Enrichment System

In District R, a program of heterogeneous grouping plus enrichment was employed at all grade levels. At the elementary school level, pupils were randomly assigned to the various sections within a given grade level. All principals in District R were briefed on the research project by the project director during the pilot study year of the project and the superintendent set up a specific policy for placing pupils in sections that required pupils to be randomly assigned. At the secondary level, the same policy was in effect.

Reduction in Heterogeneity

Because ability grouping systems attempt to make adaptation for individual differences easier by reduction of heterogeneity within the classroom, it is important in the appraisal of such a system to determine just how much of a reduction in heterogeneity is achieved. The amount that heterogeneity can be reduced by ability grouping is dependent upon several variables. Among these are the specificity and validity of the grouping criterion, the number of grouping levels to be established and the variability and numbers of pupils to be grouped. If we have large numbers of heterogeneous pupils, employ multilevel grouping and have a highly valid criterion, we can achieve great reductions in heterogeneity in a specific subject area.

A criticism often leveled at ability grouping programs, particularly those employing general criteria such as an intelligence test score or a composite achievement test score, has been that such a program reduces the heterogeneity of pupils in specific subjects to such a slight degree that the teacher, in effect, has just as heterogeneous a class in specific subject areas as would be found if no grouping had taken place.

During the pilot study year, our samples were studied carefully to determine the reduction in heterogeneity achieved by the grouping system used in District A. In order to determine the variability of those classrooms included in the research from District A and District R, means and standard deviations were computed for each elementary school classroom on the California Mental Maturity Short Form Language Score, the CAT Arithmetic total, CAT Language total, and the CAT Reading total. The average variability in District A on the various subtests was found to be 26 percent less than in District R for the elementary sample. The smallest difference in variability was found between fourth grade pupils in arithmetic where the average standard deviation was only about ten percent smaller for District A. The greatest difference was found in the reading scores of sixth grade pupils where the standard deviation was nearly 45 percent smaller in District A than in District R.

A similar overall reduction in variability was found at the junior high school level. At this level, however, the reduction in variability was usually much greater for superior and slow groups than for average groups. This is because a larger proportion of the pupils were placed in average groups at the junior high school level, and therefore, these average groups were nearly as variable as random groups in District R.

Curriculum Adjustments

We have discussed the procedures used by Districts A and R to group pupils for instruction and the differences in classroom homogeneity resulting under these systems. Grouping systems, however, are but a means to an end; this end being to provide a situation in which more adequate curriculum adjustments for individual differences are possible. Therefore, in order to evaluate the results obtained by ability grouping or enrichment programs, it is necessary that the reader have information concerning the adjustments made in these programs to adapt the curriculum to pupils at different levels of ability.

Many of the ability grouping studies carried out in the 1920's and 30's, such as Worlton's (123) research, went to great lengths to maintain identical curricula for pupils at all ability levels in ability grouped classes. In other words, every effort was made to avoid any adjustments in the curriculum that may have better adapted it to the superior or slow pupil. Such an approach, of course, overlooks the main purpose of ability grouping. Thus, in striving for better control, these studies eliminated the principal element they should have studied. Some of the studies reported in the literature have made considerable adjustments in the ability grouped classes but none in the random grouped classrooms used for comparison. This also appears to be a questionable approach because in effect, it involves comparing "something" with "nothing", i.e., a definite program versus no program to adjust the curriculum to pupil ability.

In the research reported herein, both of the districts involved had positive programs and were closely comparable in the resources available for implementing these programs. Both the acceleration employed in District A and the enrichment program employed in District R attempted curriculum adjustment and, for that matter, the two districts sometimes employed identical materials for making adjustments for different ability levels. These adjustments in District R, however, were made in classes in which no attempt had been made to reduce the heterogeneity of the group.

In general, the goal of the District A program at both elementary and secondary level was to adapt curricular materials to the different ability levels and concurrently adjust the rate of presentation to the level of the child. Thus, at a given grade level the slow group would use curricular materials appropriate to their ability and would be working on learning objectives that had probably been attained by the average groups earlier in the year and by the superior groups during the preceding year. Adjustments in rate rather than depth to provide for individual differences is characteristic of acceleration programs as opposed to enrichment programs. The curriculum that progresses at different rates for different pupils requires a careful integration from grade to grade and between the elementary and secondary programs. Otherwise the bright child will find himself repeating in seventh grade material he learned in the fifth grade and the slow child will find himself expected to have knowledge and skills that he has never been taught. Perhaps the main advantage of the accelerated program is that it permits the superior student to move to more advanced course work than is possible in an enrichment program. For example, the pupil who finishes the sequence of general science courses a year earlier than usual can take one or two more of the advanced electives in the physical sciences or natural sciences than he could otherwise fit into his high school course of study.

In contrast to District A where the main goal was to adjust rate of learning to individual differences, District R employed enrichment in heterogeneous classrooms to adjust the depth of learning to individual differences. The superior pupil in District R would therefore not take specific courses, such as algebra, any earlier in his program than would other pupils. If successful, however, such an enrichment program should develop a fuller understanding of the work covered by the superior pupil. The enrichment program also provides the bright pupil with more time for independent thought and study than would be available to him in an accelerated program. Whether such pupils generally profit from this free time or instead become lazy and develop poor study habits and attitudes is one of the questions with which the Utah study was concerned.

Let us now discuss some of the specific adjustments made by the two districts in order to provide the reader with greater insight into the setting in which this research was conducted and the success of the two systems in carrying out adjustments dictated by their respective policies of enrichment and acceleration.

District A Elementary Classes: No regular inservice training program was carried out to prepare District A teachers in the elementary grades to handle classes of either superior or low ability pupils. However, supervisors worked with the teachers and held study groups and conferences during the course of the study to help teachers improve their adaptation of the curriculum to these groups.

Adjustments in the reading program included adopting different textbooks which were appropriate in terms of reading level for the pupils in the superior, average, and slow groups at each grade level. When a child was encountered who required materials outside of the range provided to classrooms at his grade and ability level, additional materials were obtained from the supervisor on an individual basis.

The arithmetic program was also adapted to the three ability levels. This adaptation involved the use of simpler materials and a slower rate of progress at the slow ability level and accelerated progress through more complex materials at the superior level. In the area of arithmetic relatively little supplemental material was made available to the classroom teacher as the primary goal of the program was completion of the regular sequence in arithmetic at either a faster or slower rate for pupils of different abilities.

In the elementary science program, slow pupils used a textbook that was designed for the grade level previous to the one in which the pupils were enrolled. Regular and superior pupils used the regular textbook designed for their grade level. Supplemental materials for the slow classes included the basic science education series published by Rowe-Peterson for the primary grades. Superior pupils had available the same series for the junior high-school grades, while regular pupils used the intermediate grade level series. Science was taught in the District A elementary schools using predominantly the whole class instructional approach, although pupils also carried out individual projects.

The adaptation of the social studies curriculum for superior pupils was made primarily through the use of supplemental materials. Relatively little adjustment was made in the social studies curriculum for slow pupils.

District R Elementary Classes: Classroom libraries were provided in all District R elementary classrooms. These varied in size but were generally quite extensive, and contained enrichment materials aimed at adapting the curriculum to individual differences. Most of the teacher guides used by elementary school teachers in District R suggested materials and methods for adjusting the curriculum for different ability levels within the classroom.

Perhaps the most extensive adjustments in the District R elementary classes were in the teaching of reading. Reading materials over a wide range of levels were available in each classroom. Intra-class grouping was used for reading instruction.

Fewer materials for adjusting the District R curriculum to the different ability levels were available in arithmetic than in the other subject areas considered in this research. Arithmetic teaching was carried out on a whole class basis and the time available to the teacher for working with individual children was limited.

The science curriculum in District R called for whole class instruction coupled with individual projects. Extensive enrichment materials were available in each classroom for adapting the curriculum to both slow and superior pupils although more materials were available for superior pupils than for slow pupils.

The social studies curriculum at each grade level in District R employed a single textbook for all pupils with a moderate amount of supplemental materials in the classroom libraries.

Curriculum Adjustments in the Secondary Schools: At the junior high level, District A adjustments for the different ability groups in mathematics were achieved primarily through adjusting the rate at which pupils covered the mathematics curriculum. Superior pupils covered about four years of work during the three year junior high school period, average pupils covered the usual three year sequence and slow pupils covered what would be two years of work for the average group.

In making adjustments to the science curriculum, District A used both enrichment materials and adjustments in the rate at which these materials were presented. Again, the superior pupils covered work equivalent to that usually covered in four years by average pupils and also worked to a limited degree with advanced enrichment materials designed to build greater depth. Average pupils in science followed the usual junior high school science sequence while slow pupils covered the equivalent of two years of the average science program during their three years and worked extensively with supplemental materials designed for slow learning pupils.

In District R, a wide selection of enrichment materials was available in junior high school mathematics classes. In science, a very large selection of enrichment materials was available; these materials having been purchased for the most part through an NDEA grant. Science teachers had carried out surveys to identify enrichment materials to be purchased and were therefore familiar with the materials available. More enrichment materials were available for superior pupils than for slow pupils. At the junior high school level many of the materials used in District R for adapting the curriculum in heterogeneous classes through enrichment were also used in the ability grouped classes of District A. Both districts used the Rowe- Peterson Enrichment Series, for example.

At the high school level, both districts required only one year each of mathematics and science. The electives taken by pupils thus became difficult to control and differences between the ability grouped and random grouped treatments became increasingly smaller. Ability plays a large role in all secondary schools in determining which students will take the more advanced electives in science and mathematics.

Thus, the treatment differences in the Utah study were substantial at the elementary level and it is felt the findings can be accepted with considerable confidence. The treatment differences at the junior high school level, although subject to some limitations in the ninth grade, were still large enough so that comparisons between the two treatments are felt to be reasonably valid and well founded and should have many implications for teachers and administrators. At the high school level, however, the findings should be regarded at best as tentative. Results for the tenth grade are probably meaningful indications of treatment differences because fewer electives were available during the tenth grade and furthermore, tenth grade results can be regarded to a considerable degree as reflecting the effects of earlier work at the junior high school level. At the eleventh and twelfth grades, however, a wide range of electives is available and there is a tendency for ability grouping to occur naturally in both districts because of the restricted ability range of pupils choosing mathematics and science electives. These circumstances lead the principal investigator to suggest that treatment differences are limited at this level and the reader should interpret the findings with caution.

CHAPTER 2

ACHIEVEMENT-RELATED VARIABLES

THE GOAL of this section is to explore the relative effectiveness of ability grouping and random grouping in terms of achievement-related performance of pupils under these two treatments. We will examine the rather extensive previous research that has been carried out in this area, and will then add the findings of this project. In the Utah study, achievement data were collected for all subjects at the beginning of the study, and on at least three additional occasions during the subsequent four years. Achievement data were collected on at least one of the five samples at every grade level from fourth through twelfth, and for some grade levels, data were collected on as many as four different samples. In addition, studies of overachievers and underachievers and study methods were carried out.

PREVIOUS RESEARCH ON ACHIEVEMENT AND ABILITY GROUPING

General Achievement

Over the years, perhaps the major appeal of ability grouping to the educator has been its promise to provide a classroom situation in which better adjustments of the curriculum for individual differences might be possible. The goal of such curricular adjustments generally is to bring about higher achievement by providing more adequately for individual differences. In view of this appeal, it is not surprising to discover upon reviewing the research literature related to ability grouping that the vast majority of studies in this area are concerned with achievement gains in ability grouped and random grouped classrooms.

A number of reviews of research in this field have been published. The reader is recommended to Ekstrom (34) who has done a thorough job of reporting and critically evaluating the research on ability grouping up to the date of her review. Rather than duplicate this and other excellent reviews that have been published, we will review in detail only a few outstanding studies, many of which have been completed since Ekstrom's review was published. This rather detailed coverage has been supplemented by two tables that give the essential information on a large number of additional studies that may interest the reader. Table 2 covers studies that have been carried out at the elementary school level, while Table 3 deals with studies at the secondary school level. Even these tables are somewhat selective as they do not include many of the weaker experimental studies nor any of those publications that present opinions or personal experience rather than research evidence.

In our discussion, we will point out a number of weaknesses and limitations. These criticisms are not made with the intention of minimizing the importance of the knowledge that these studies have given us, but instead attempt to provide the reader who is unfamiliar with this research area with some insights that will help him place the various studies in their proper perspective.

Elementary School Level: Although completed 30 years ago, Hartill's research (58) is probably the best designed work published to date on the relative merits of homogeneous and heterogeneous grouping at the elementary school level. His subjects included 1374 pupils enrolled in grades 5A, 5B, and 6A. On the basis of educational age, one-half of the subjects were grouped homogeneously into three ability levels while the other half were assigned to heterogeneous groups. Pupils in the homogeneous groups received a differentiated course of study aimed at their ability level, while those in the heterogeneous groups received an undifferentiated course of study. After the subjects had been exposed to one-half year in the homogeneous or heterogeneous classes, they were again tested for achievement.

The following fall, when these pupils returned to school, those who had been in homogeneous groups during the previous school term were placed in heterogeneous groups, and vice versa. Teachers were also advanced so that in a given school the teachers who had taught 5A children who were grouped homogeneously would teach the same children in grade 5B when they were placed in heterogeneous groups. At the end of the fall term, all pupils were again tested for achievement. An overall analysis of the achievement results for pupils when in homogeneous and heterogeneous groups showed no significant differences between the gains made by these pupils under the two conditions. When pupils were subdivided in terms of ability level, however, it was found that the superior pupils made larger gains in achievement areas tested (except reading) when under the heterogeneous grouping system. Average pupils made larger gains when under the homogeneous grouping system, and slow pupils did equally well under the two systems. It should be noted that superior pupils in the homogeneous groups were

TABLE 2

ABILITY GROUPING STUDIES CONCERNED WITH ACHIEVEMENT AT THE ELEMENTARY SCHOOL LEVEL

Study	Grade Levels	N	Length of Study	Grouping Levels	Grouping Criteria	School Subjects Measured	Results	Remarks
Jones, McCall (1926)[a]	3-7	67 Pairs	2 years	Superior only	IQ	Reading Arithmetic Spelling	Slight advantage for AG sup. groups, not sig.[b]	AG pupils elected to enter accelerated classes, thus biasing sample. Matched pairs used. Curriculum differentiated for AG pupils.
Worlton (1926)	4-7	3700 AG and RG	1/2	3	IQ	Reading Arithmetic Spelling	AG classes made greater gains in composite achieve. score (no mention of stat. sig.)	Steps taken to avoid differentiation of curriculum for different ability levels.
Baird (1927)[a]	4-8	325 in AG + 10 classes RG		3	Ach. Test	Gen. Ach.	8 of 10 control classes achieved more (no mention of stat. sig.)	
Walton (1928)	4-7	1466		IQ Levels	Teacher Judgment and IQ	Reading Spelling Arithmetic	Slight advantage for AG sup. groups on composite achieve. score (no mention of stat. sig.)	1466 pupils above 90 IQ were tested. Pupils were classified into 6 categories each with a range of 10 IQ points. Curriculum differentiation implied but type of adjustments not described.
Bonar (1929)	1	96	1 year	2	IQ and Teacher Judgment	Reading	Results on objective achieve. test favored RG. No data on stat. sig.	The grouping system used did not result in really different ability groups. The median IQ of the high group was 104, the low group 98 and the het. group 99.
Dvorak, Rae (1929)	1	20 Pairs	1 year	2	IQ	Reading Spelling	No clear-cut dif. bet. AG and RG. Two stat. sig. dif. in reading scores favored AG.	The grouping system used could not be expected to reduce the variability of the class by more than a few percent.
Gray, Hollingsworth (1931)	Age 7-9	56 exp. 36 control	3 years	Superior only	IQ (S.B) Above 130	Reading Arithmetic Nature History Language	No stat. sig. dif. found in regular school subjects. "Enrichment" subjects studied by AG pupils not measured.	AG pupils did extra work in some subjects and studied some topics not covered by RG pupils.
Barthelmess, Boyer (1932)	4-5	297 Pairs and 268 Pairs	1 year	3	IQ	English Reading Geography Arithmetic	Total AG sample was sig. higher on total achieve. for both replications. AG pupils at all 3 ability levels made greater gains but sig. level of these gains not reported.	Study involved 2 replications. Used matched pairs which resulted in loss of up to 53% of AG pupils and 77% of RG pupils.
Hartill (1936)	5-6	1374	1 year	3	Ach. test	Gen. Ach.	Overall comparisons bet. AG and RG pupils showed no stat. sig. dif. Sup. RG pupils and av. AG pupils made greater gains. No dif. between slow groups.	Well controlled study using counterbalanced design, to control group dif. Differentiated course of study used with AG sup. pupils being given enrichment program rather than acceleration.
Rankin, et. al. (1936)	3-6	About 500 in each plan	2 years	3	IQ	Reading Arithmetic	On composite score of 4 reading tests, vertical plan best at all ability levels. Detroit plan higher than RG for composite and bright pupils but these 2 plans about same for av. and slow pupils. In arithmetic Detroit plan best, vertical plan next and RG lowest for all ability levels. Dif. greatest for sup., smallest for slow groups.	Two dif. AG plans compared with RG plan. Vertical plan cut across grade levels to form AG; Detroit plan grouped within grade levels.
Kvaraceus, Wiles (1938)	3	75	1 year	3	Ach. test	Reading English Arithmetic	"Better than av. pupil growth", no mention of stat. sig.	Pupils grouped separately in reading, English, and Arithmetic and grouped het. rest of day. Major effort made to differentiate curriculum. No control group.

TABLE 2 (continued)

Study	Grade Levels	N	Length of Study	Grouping Levels	Grouping Criteria	School Subjects Measured	Results	Remarks
Russell (1946)	4-6	278 exp. 248 control	2 years	3	Reading Ach.	Reading	No stat. sig. dif. in gains for either the entire groups or for pupils at dif. ability levels.	Involves AG in reading only with pupils in het. groups for rest of school day. A well designed study.
Edmiston, et. al. (1949)	5-6	462	1 year	See Remarks	IQ Range	Reading	Wide range group made sig. greater reading gains.	Compared "wide range" groups (IQ Range 41) and "narrow range" groups (IQ range 29). Curriculum not differentiated.
Husen, Svensson (1960)	5-6	2755	2 years	2	School marks based on Standard Test	Reading Grammar Spelling Arithmetic	As analysis was broken down by treatment, social class, and ability level, a brief summary of results is not possible. Few sig. dif. found bet. treatments for pupils of comparable ability.	This study was carried out in Sweden.
Provust (1960)	4-6	254	1 year	3	IQ	Arithmetic	Overall comparison gave sig. greater gain for AG. Analysis by level showed AG sup. made greater gains. No dif. bet. AG and RG av. or slow pupils.	Experimental group in AG class for arithmetic only; in RG class for rest of day. Study used matched pairs.
West, Sievers (1960)	5-6	16 exp. 16 control		Superior only	Teacher Judgment and Ability Test	Language Art and Arithmetic	Mean gains of 2 to 5 months for AG over RG pupils.	Matched pairs design. Use of teacher judgment probably introduced bias favoring experimental group. Experimental group was in special AG class half day and RG class remainder of day.
Daniels (1961)	Junior School 1-4 yrs.	About 500 cases. Varies from year to year.	4 years	3	IQ and previous grades	English Reading Arithmetic	16 dif. bet. means for one pair of schools, 4 for another pair of schools. All 20 reported sig. favoring RG schools. No breakdown of data by ability level.	No data given on curriculum differentiation in AG schools. Groups matched for mean and SD of IQ. Author apparently used one-tailed test of sig. as CR's as low as 1.75 are reported sig. at .05 level.
Goldberg, et. al. (1961)	4	2219	1 year	5 See Remarks	IQ	Reading Arithmetic Language Arts, Work Study	Of 105 comparisons, there were 11 that were stat. sig. 10 of these favored "broad range" patterns i. e. het. grouping.	5 ability levels combined into classes to provide 15 patterns differing in ability levels and degree of hom. Curriculum not differentiated.
Koontz (1961)	4	191 to 231 for dif. subjects	1 year	5	Arithmetic Reading Language	Arithmetic Reading Language	RG sample achieved sig. more in reading and arithmetic. No sig. dif. in language. No sig. interaction bet. treatment and level.	AG and RG pupils matched on initial achieve. Analysis of variance used to analyze results. Hawthorne effect probably favored AG.
Simpson, Martinson (1961)	5-6	295 in grades 5-6	1 year	Superior only.	IQ (SB) Above 130	Arithmetic Reading Writing Listening Social Studies	Both the special enrichment group and special hom. group achieved sig. more than control group on gen. ach.	This study involves a wide variety of treatments for sup. pupils in grades 1-12. All treatments are compared with control groups but treatments are not compared with each other.

---

[a] Taken from Ekstrom, (1959). Primary source not available.

[b] Abbreviations:

AG = ability grouped
RG = random grouped
stat. = statistical, statistically
sig. = significance, significant
sup. = superior
av. = average
dif. = difference(s), different
hom. = homogeneous
het. = heterogeneous
ach. = achievement (also achieve.)
bet. = between
gen. = general

TABLE 3

ABILITY GROUPING STUDIES CONCERNED WITH ACHIEVEMENT AT THE SECONDARY SCHOOL LEVEL

Study	Grade Levels	N	Length of Study	Grouping Levels	Grouping Criteria	School Subjects Measured	Results	Remarks
Theissen (1922)	7	65 to 74 classes	1 sem.	6	IQ	Arithmetic Reading Language	Compares AG sections at dif. ability levels rather than comparing AG and RG classes. The poorest of the top IQ classes achieved less in both arithmetic and language than the best of the bottom IQ classes.	Demonstrates the importance of teacher dif.
Cook (1924)	9-10	495	1 sem.	2	Past Grades	English History Geometry	No dif. bet. AG and RG in English. Sup. AG better in history, poorer in geometry. Slow AG poorer in history, better in geometry.	
Moyer (1924)	9		1 sem.	3	IQ	Algebra Latin	Sup. AG better in Latin. Av. AG better in both subjects. No dif. bet. slow AG and RG.	No data on stat. sig.
Purdom (1924)[a] and (1929)	9		1 sem.	3	IQ	English Algebra	No dif. bet. AG and RG.	
Woody (1924)	9	51 Matched Pairs	1 sem.	3	IQ	English	No dif. bet. AG and RG. Results not broken down by ability level.	Used matched pairs.
Martin (1927)[a]	7		1 year	3	IQ	General Achieve.	No sig. dif. bet. AG and RG on general achievement test.	Used matched pairs. Teachers told not to differentiate the curriculum.
Van Wagenen (1927)[a]	7-8	188		Superior Only	M. A.	General Achieve.	No sig. dif.	
Billett (1928)[a]	9		3 years	3	IQ	English	No sig. dif. Largest dif. favored AG slow pupils.	
Holy, Sutton (1930)[a]	9		1 sem.			Algebra	AG made greater gains.	
Fowlkes (1931)	7				IQ	General Achieve.	AG gained slightly less than RG on general achieve. test.	

Rankin (1931)	7	730	1 year	3	C. A. IQ and Teacher Judgment	English Arithmetic Gen. Sci. Soc. Sci.	Slow and Av. achieved more in AG. Sup. achieved more in RG. No dif. in results for dif. school subjects.	
Herr (1937)	7-8	97	1 year	Superior Only	Past achieve. IQ and Teacher Judgment	English History Plane Geometry, Chemistry and general information.	Dif. in American History and general information favored AG.	Matched pairs design. AG accelerated to cover 2 years work in 1.
Justman (1954)	7-9	79 to 83 Matched Pairs	2 years	Superior Only	IQ Above 130	Math. Science Soc. Stud.	AG pupils achieved sig. more at .01 level in all 3 school sub. at end of study. No evidence that groups were comparable at beginning of study.	Matched pairs design. This study is biased in favor of AG as highly motivated pupils selected for AG leaving less motivated pupils in RG. AG completed 3 years of Jr. H. S. in 2 years.
Wilson (1959)	9	96 Pairs Superior 48 Pairs Average	1 year	2 Superior and Av.	IQ and Teacher Judgment	Soc. Stud. Reading Writing Use of Sources, Problem Solving, Critical Thinking.	There were not stat. sig. dif. in gains made by pupils in AG and RG classes.	Pupils were grouped in special combined English-Social-Studies classes only. Sup. group IQ's above 115; av. group 100-110; no slow group studied.
Passow, Goldberg, and Link (1961)	7-9	112	3 years	Superior Only	IQ	Math.	Accelerated gifted pupils achieve. sig. more in math. than enriched gifted pupils or control group.	Used 4 matched homogeneous groups of gifted pupils. Tested effectiveness of dif. cur. adjust. for homogeneous gifted groups.
Simpson, Martinson (1961)	8 11-12	436 (sec.)	1 year	Superior Only	IQ (S. B.) Above 130	Arith. Reading Writing Listening Science Soc. Stud.	Special. hom. classes at each grade level achieve. sig. more than RG controls.	This study involves a wide variety of treatments for sup. pupils in grades 1-12.
Drews (1962)	9	432	1 year	3	IQ Reading Skills, Language Skills	Problem Solving, Reading, Language, Critical Thinking	No stat. sig. dif. bet. AG and RG pupils on any of the ach. measures.	Carefully designed and executed. The only previous study which compares programs that provide extensive differentation of the curriculum for both AG and RG pupils.

---

[a] Taken from Ekstrom (1959).

Abbreviations:
AG = ability grouped
stat. = statistical, statistically
sup. = superior
dif. = difference(s), different
bet. = between
RG = random grouped
sig. = significance, significant
av. = average
hom. = homogeneous
ach. = achievement (also achieve.)

given an enrichment program rather than an acceleration program, and were, in fact, held back to a normal degree of growth on the fundamentals so as to permit more time for enrichment. Hartill concluded that the homogeneous plan was superior because bright pupils did nearly as well on the fundamentals and in addition had time to study the enrichment materials, average pupils gained significantly more, and slow pupils did just as well under this system.

Hartill's research is the only study found in the area of ability grouping in which a counterbalanced research design has been used. This design has the great advantage of cancelling out differences between experimental and control groups because each pupil is measured in both treatments during the two phases of the study. Hartill's design also controlled teacher differences, by having the same teachers move ahead one-half grade so that within a given school, the same teachers would teach both homogeneous and heterogeneous classes.

Another of the early studies that involves large samples and gets at basic questions with a reasonably sound research design was carried out by Rankin, et al., (93). This study involved about 500 pupils in each of three grouping plans, and compared achievement gains on four reading tests and three arithmetic tests over a two-year period. The three grouping plans used were: 1) the "vertical plan" in which pupils were grouped on three ability levels combining as many as three half-grades to make up a class, 2) the "Detroit plan" which also employed three ability levels, but did not cut across grade lines, and 3) the "mass instruction plan" in which no attempt was made to establish homogeneous groups. Curriculum adjustments were made in both of the ability grouping plans although these adjustments were greater in arithmetic under the Detroit plan than under the vertical plan. In the mass instruction plan "... a deliberate attempt was made to teach all pupils in the same way. This was done by using the same material and the same methods for all and attempting to hold all pupils to the same standards. Standard Detroit materials were replaced in these schools by other materials better adapted to mass instruction." (p. 279).

The results of this research are given in terms of gain scores over a two-year period. Gains are given for all pupils on each of the tests and are also presented for bright, average, and slow groups separately. Although the research report presented no data on statistical significance, the information provided is sufficient for calculation of t-tests between some of the final means. These t-tests were calculated for the Stanford Reading Test, which was weighted about half of the composite reading score. The results on the Stanford Reading Test comparing all pupils regardless of ability levels showed the vertical plan to be superior to both the Detroit Plan and mass instruction plan with the differences significant beyond the .01 level. Differences between the Detroit and mass instruction plan were not statistically significant. When comparing bright pupils only on the Stanford Reading Test, the results again showed that the vertical plan resulted in greater gains than the other two plans, with these differences also significant at the .01 level. The pattern was somewhat different, however, for the slow group. Slow pupil gains under the vertical plan were significantly higher at the .05 level than under the Detroit plan, but were not significantly different from the mass instruction plan. Again, the Detroit and mass instruction plans were not significantly different from each other. It is interesting to note that there were no significant differences between heterogeneous grouping (mass instruction) and homogeneous grouping within grade level (Detroit plan). Those significant differences found favored the more extreme system of homogeneous grouping across grade levels that was employed in the vertical plan. T-tests were also calculated by the reviewer on the Detroit Arithmetic Test which was weighted about one-half of the composite arithmetic score. None of the differences between the three grouping treatments on arithmetic gains were statistically significant.

Probably the greatest limitation in the findings of this study was brought about by the removal of all enrichment materials, even those normally used by the district, from the heterogeneous classrooms involved in the "mass instruction plan". In taking this step, the investigators created an unusual difference between the treatments in the amount of supplemental materials available, and as a result, the achievement differences obtained can be considered only partially attributable to the grouping treatments. The fact that reading differences were much larger than arithmetic differences suggests that the absence of material for differentiating the curriculum might have had greater influence on the results than the grouping treatment, because the teaching of reading without any materials available for adapting to individual differences presents the teacher with a far more difficult task than teaching arithmetic under these conditions. In spite of this limitation, however, this is one of the best of the comparative studies because it involves reasonably large samples, uses reliable objective measures, breaks down the effect of the treatments for pupils of different ability levels, and avoids the pitfalls of using matched pairs of subjects.

Among the more recent studies, the work of Daniels (29) is one of the most extensive at the elementary level. In this study of four British junior schools (grades 2-5 in our system), two employed three-level ability grouping based on IQ and two employed random grouping. Although little information is provided on the extent of curriculum adjustment in the ability grouped schools, the report indicated that the four schools had comparable curricula and the random grouped schools used whole group instruction rather than enrichment or intraclass grouping. Schools A and B employed random grouping while schools C and D employed ability grouping or "streaming" as the English call it. Schools A and C were found to be comparable on a number of pertinent variables as were schools B and D. Group matching procedures were used to equate the groups for mean and standard deviation of IQ. This matching resulted in losing relatively few cases. Three achievement areas; reading, English and arithmetic, were tested at the start of the study and at intervals during the study. For schools A and C, the combined number of cases for different years ranged from 262 to 297. For schools B and D the combined N ranged from 197 to 227.

In comparing mean achievement differences in the three subject areas between school A and school C, Daniels reported all 16 critical ratios between the two treatments at various intervals ranging from a half year to four years to be statistically significant. All of these differences between means were in favor of school A, the school that employed random grouping. Only four comparisons between schools B and D are reported in the achievement areas but all four of the critical ratios obtained were statistically significant favoring school B, the random grouped school. All data were analyzed for total groups rather than being broken down by ability levels.

Daniels used the rather unwieldy procedure of rematching his groups at various intervals in order to measure gains at different periods over the four years. For example, to measure the significance of gains during the third and fourth years in English, he rematched his groups on the English test scores given in the second year and then compared the mean English test scores that the groups so matched earned during the fourth year. Analysis of covariance would have been a far simpler and more effective tool for this study. In general, however, Daniel's study appears well designed and carefully carried out. If the data had been analyzed separately for pupils at different ability levels, this study would have provided a great deal more useful information.

Another recent study by Goldberg, et al., (46), is by far the most extensive project to date aimed at evaluation of ability grouping per se, that is, without any attempt at adjustment of the curriculum for groups of different ability levels or degrees of homogeneity. This research was carried out in the New York Public Schools starting in spring of 1956. Subjects were selected from the fourth grade level. Five ability levels were identified, ranging from Group A with IQs of 130 or higher to Group E having IQs of 99 and lower. Classes were organized so as to contain different combinations of the various ability levels. For example, Pattern 1 contained only pupils with IQs of 130 and above (Group A); Pattern 2 included pupils from both groups A and B; Pattern 3 included pupils from groups A, B and C and so on. A total of 86 classes were formed which were classified into 15 patterns representing different combinations of pupils at the five IQ levels. A wide range of measures were administered to the sample including academic achievement in reading, arithmetic, language arts, work study skills as measured by the SRA Achievement Series, and science and social studies as measured by the Stanford Achievement Tests. Findings of this study in general indicate that when no attempts are made to adjust the curriculum for different ability levels, the process of narrowing the ability range in the classroom is unlikely to raise pupil achievement. In this research, as a matter of fact, Pattern 5, which represented the broadest ability spread among the 15 patterns, appeared to be most consistently associated with greater academic gains for all pupils. The achievement differences between patterns, however, tended to be small with no one pattern or combination of patterns resulting in better achievement for all pupils in all subject areas.

The findings of this study indicate that achievement was influenced more strongly by the teacher and group differences in individual classrooms than by the ability range of the class. Advocates of heterogeneous grouping have often argued that the presence of superior pupils tends to lead to greater achievement by pupils of lower ability. This contention was tested in Goldberg's study and it was found that only in science was the presence of gifted pupils consistently related to greater increments for the other four ability levels. In other school subjects measured, the presence of gifted pupils was not related to greater achievement gains for non-gifted pupils.

A further analysis of the data was accomplished by combining the 15 ability patterns into three broad categories: narrow range, medium range, and broad range classes. An analysis of variance indicated that the ability range in the classroom had a significant bearing on the achievement in social studies, reading comprehension, vocabulary, arithmetic, and total average achievement; while science, language arts, and work study skills were unaffected by the range. In the six academic areas affected, overall comparisons showed greater gains for pupils in the broad range classes as compared with the medium or narrow range patterns. When each ability level was taken separately, the effects of range were inconsistent, although revealing some significant differences. Of the 105 possible comparisons between pupils in the five ability levels, the three range patterns and the seven academic areas, only 11 reached significance, and in 10 of these the broad range was superior to either or both of the other two ranges.

This study also explored the question of whether the pupil's position in his group in terms of ability was related to his achievement. This is an important question because one of the commonly used arguments against ability grouping is that it deprives the slow pupils of the chance to learn from more able classmates. The analysis showed that in most subjects the presence of one or two levels of more able classmates and no slower ones did not have any positive effect upon achievement.

The research carried out by Goldberg and her associates is very well designed and involved a thorough and sophisticated analysis that deals with most major questions concerned with class ability level and degree of heterogeneity as it relates to achievement.

When the overall results of the study are considered, one must conclude that variations in the ability level, the ability range, and the position of individual pupils within the ability range in classes where no specific curricular adaptations are made, has little effect upon pupil achievement. In other words, ability grouping per se offers no achievement advantages over random grouping and appeared less effective than broad range grouping under the conditions present in Goldberg's study. This is one of the very few areas related to ability grouping in which scientific evidence from an extensive and well-designed study seems to have presented a clear and reliable answer.

Secondary School Level: Nearly all studies concerned with ability grouping versus random grouping at the secondary level have deficiencies that raise serious questions about the validity of their findings. We will limit this review to a few of the recent studies that appear to provide reasonably valid research evidence. The reader is referred to Table 3 for a brief summary of the major studies concerned with ability grouping at the secondary school level.

Starting in 1956, an experimental four-track curriculum was established in the Washington, D. C. high schools. The following year this program was extended to the eleventh grade and in 1958 to the twelfth grade. When the first class progressing through the entire sequence had graduated, an evaluation of the program was carried out and reported (57). This four-track program was essentially an ability grouping program designed to reduce the ability range of each classroom and increase efficiency of instruction by differentiating the curriculum. The four tracks included the honors track that was restricted to pupils above 115 IQ, the college preparatory track, the general track, and the basic track. The honors track was subjected to an extensive comparative evaluation in which seniors from the class of 1958 who were not in the track program were compared with seniors from the class of 1959 who were in the honors track during their high school years. The study involved the top 270 graduates of each year, the groups being selected so as to give an even proportion of boys and girls and very similar composition in terms of IQ. Both groups had been administered the Iowa Tests of Educational Development during their senior year. The median score for the honors track students of 1959 was five points higher than the comparison group of 1958 in "correctness of expression", three points higher in "quantitative thinking", and six points higher in "science achievement". The two groups were also compared on the College Entrance Board Examination. The honors program group was 37 points higher on their median scholastic aptitude score, and 30 points higher on their median mathematics score. Although the research report does not give levels of statistical significance for these differences, it appears likely that all of the differences reported are significant. This study presents rather sound evidence in favor of a highly differentiated curriculum for superior students. A thorough analysis of the other three tracks as they compare with heterogeneous grouping would have added greatly to the value of the study.

Probably the best controlled and most informative study concerned with ability grouping at the secondary level that has been reported to date is the research recently completed by Drews (31). Although limited in some respects in that pupils were studied at only one grade level, and in one subject area, this research used exceptionally good design and assembled data on a number of important variables that have not been explored in previous research. The goal of Drews' research was to determine the effects of heterogeneous and homogeneous grouping upon pupils at three ability levels in ninth grade English. The criteria for establishing the pupil's ability level were IQ, reading skills, and language skills. Placement of pupils into the two treatments and the three ability levels was carried out using stratified random sampling prior to the start of the school year. Tests and questionnaire data were collected at the beginning and end of the year, and data were also gathered in observations, three being made of each class involved in the study. The final sample for whom all major data were collected included a total of 432 students. Because both the homogeneous and heterogeneous groups were divided into three ability levels, and each ability level was divided by sex for the purposes of analysis, some of the sub-samples were small.

Eight teachers were involved in the experiment: four experienced and four inexperienced. As a rule, each teacher taught two homogeneous classes at different levels and one heterogeneous class, thus helping to cancel out possible differences in teaching ability. Experienced and inexperienced teachers taught equal proportions of the homogeneous and heterogeneous classes. All teachers met with Drews for a 90 minute period every two weeks throughout most of the experimental year in an effort to develop a common viewpoint on the methods and curriculum adjustments to be made in the homogeneous and heterogeneous classes. Teachers attempted to adapt the curriculum in both homogeneous and heterogeneous classes. Some acceleration was carried out in the superior homogeneous classes, but the main goal of the differentiated curriculum at all ability levels was enrichment. One of the outstanding aspects of Drew's study was the great effort aimed at helping teachers develop a well-differentiated curriculum for use in both homogeneous and heterogeneous classrooms.

Initial and final achievement were measured in Drews' study using the reading and language tests from the California Achievement Test Battery. The advanced form was used for superior pupils, the intermediate form for average pupils, and the elementary form for slow pupils. In addition to the reading and language tests given at all ability levels, superior students were also administered tests of critical thinking and dogmatism and average pupils were administered tests of problem solving and dogmatism. The slow group was given an individual reading test in addition to the group tests of reading and language. Initial and final test scores were analyzed using the t-test to compare homogeneous and heterogeneous sub-samples of the same sex and ability levels. In reading comprehension, none of the comparable groups were significantly different on either the initial or final testing. The same comparisons of the pre- and post-tests of language achievement yielded the same results, although some of the differences obtained approached statistical significance at the .05 level. Negative results were also obtained on the critical thinking tests, problem solving tests, and dogmatism scale. Thus, it may be concluded from Drews' study that homogeneous grouping in ninth grade English has no significant effect upon the achievement areas measured at any of the three ability levels. In addition to the achievement data given above, Drews' study also dealt with a number of other important variables that will be discussed in the non-cognitive section of this report.

Drews' study, although very well designed, had two limitations worthy of mention. One of these, brought out by Drews herself, was that the achievement measures used were aimed primarily at measurement of facts, while the type of program that Drews was attempting to stimulate might be expected to develop many aspects of achievement not measured by a fact-oriented test. The study collected other data that were aimed at measurement of these higher goals, but this weakness could have been further overcome by using an achievement battery such as the Sequential Tests of Educational Progress, which places more emphasis upon higher cognitive processes than does the California Achievement Test.

Another minor weakness of Drews' study was her use of the t-test for measuring both pre- and post-test differences, rather than the use of analysis covariance which would have adjusted the post-test mean scores for pre-test differences that she found. Although it is doubtful that the use of analysis of covariance would have made any major change in Drews' results, it would appear from studying her data that two or three of the post-test differences would have reached statistical significance if this analysis technique had been employed. Even considering these limitations, Drews' study is exceptionally well done and probably provides more sound scientific evidence on the effects of ability grouping at the secondary level than all previous research combined.

Critical Evaluation

A critical evaluation of studies concerned with ability and achievement reveals a number of weaknesses. Perhaps the most serious of these is that many studies are of only one year duration. In most of these studies the ability grouping method is the "new" or "experimental" treatment. This situation tends to lead to a significant Hawthorne Effect which may well explain all or most of the differences found in these studies.

Another weakness of many of these studies is the use of inadequate analysis techniques. Many of the early studies and some of the more recent ones report no tests of statistical significance but merely indicate that one or the other group did "better" or "achieved more" (85, 123, 63, 12, 96).

Billett (10), in one of the early reviews of research in this field indicated that, of the studies reported between 1910 and 1928, only four made reasonable attempts to control pertinent variables. Other investigators have been content to make a single comparison between ability grouped and random grouped samples and have made no attempt to appraise possible differential effects of ability grouping versus random grouping for pupils at different ability levels (29, 121). As there is no reason to assume that ability grouping affects superior, average, and slow pupils in the same way, these overall comparisons can have little real meaning. It is at least hypothetically possible, for example, for superior pupils to make much greater gains in ability grouped classes, average pupils to do equally well in either grouping situation , and slow pupils to make much greater gains in random grouped classes. In this case, if pupils of different ability were lumped together for analysis, the lower relative gains of the slow pupils would cancel out the higher gains of the bright pupils and lead the investigator to conclude that the two grouping systems brought about no difference in achievement. Thus, such gross comparisons can give us little insight into what really happened in an experiment.

A design weakness in many ability grouping studies is the use of matched pairs of subjects (91, 66, 77). Many of the deficiencies of matching techniques such as the tendency for matched groups to be biased with respect to variables other than the matching variable, and the biases introduced by loss of cases because of non-random factors, were generally unknown when many of the earlier studies were carried out. These deficiencies, however, have been generally recognized by researchers for the past 15 years, but, nevertheless, a study reported in 1960 (90) employed matched pairs which resulted in the loss of about one-half of the cases involved in the experiment.

A number of other studies (66, 59) have used techniques for selecting pupils for ability grouped classes which bias the sample and seriously distort the findings. In a study by Justman (67) for example, pairs of intellectually gifted pupils from "special progress" and "normal progress" classes were matched on a number of variables including grade level, sex, chronological age, and mental age. Pupils for the "special progress" classes which, in this case, were homogeneous groups of superior pupils, however, had been especially selected as those "... who show superior academic achievement and who also possess personal characteristics of initiative, enthusiasm, willingness to work, reliability, regular attendance, and capacity for sustained work." (67). As all of these variables have an obvious relationship to achievement, and as the pupils assigned to homogeneous "special progress" classes were especially selected for possessing these characteristics, it is hardly a surprise to learn that these pupils made significantly higher achievement scores at the end of the study than did their equally intelligent but less motivated contemporaries, who remained in the normal progress classes. These results become more suspect when we learn that no attempt was made to compare the groups on achievement at the start of the study, at which time the "special class" group was probably already superior in the achievement areas measured. Thus, although the study was well executed except for the aforementioned design errors, it seems unwise to place any confidence in Justman's findings in view of these errors.

Westmayer (118) appears to have introduced a similar bias into his research. He compared superior pupils who elected to take a special advanced problem class in science with pupils of similarly high ability in heterogeneous science classes. Although pupils were matched on the basis of grade point average for previous high school work, it would appear that the factors that lead some pupils to volunteer for an advanced problems course are related to motivation and interest in science and therefore would be expected to be related to science achievement. The results of this study showed that the homogeneous group selecting the advanced problems class obtained significantly higher scores at the end of the study on a standardized achievement test. It is impossible to determine, however, how much of this difference was due to ability grouping, and how much was due to sampling bias. In evaluating research projects in this area, it is probably safe to assume that when pupils who elect to take a special class are compared with equally intelligent pupils who have not elected to take the class; or when pupils are placed in special homogeneous groups because they possess special characteristics not possessed by pupils in the control group, the sampling bias so introduced will account for much of the difference found.

While the aforementioned studies suffer principally because of the lack of sufficient experimental controls, the findings of some studies found in the ability grouping literature suffer from having been controlled too much. Ability grouping, as has been stated earlier, is essentially aimed at providing a classroom situation in which more adequate curriculum adjustments for individual differences may be made. Thus, studies of ability grouping that do not make some provisions for differentiating content and method for the different levels of ability cannot produce evidence on some of the major questions in this area. Several of the earlier studies including those carried out by Ullrich (115), Burtt, Chassel, and Hatch (21), and Worlton (122) attempted to keep course content and method identical for all ability levels and for both homogeneous and heterogeneous classes in their studies. These efforts at control, of course, tend to hold constant the very area that should be manipulated in the ability grouping experiment. Such studies are valuable in answering questions about ability grouping per se, but tell us nothing about the effects of ability grouping under the conditions that it is used in the field – that is, with curriculum differentiation. Some recent studies have attempted to study the merits of ability grouping per se by developing or identifying classes that vary in terms of homogeneity and ability level and then compare achievement for these classes. A study of French (37) employed this technique of locating and comparing classes already in existence which happen to differ in degree of homogeneity, while the research reported by Edmiston and Benfer (33) manipulated the degree of homogeneity while making no other adjustments. The extensive research of Goldberg, et al, (45) discussed earlier in this chapter was also aimed at evaluating ability grouping per se.

RESEARCH RELATED TO UNDERACHIEVEMENT

In addition to studying general achievement in the two grouping systems, the Utah study was concerned with the number of overachievers and underachievers in the two systems. No research has been reported that related overachievers or underachievers to ability grouping, but a number of studies on the nature of underachievement have been reported.

Some of these studies will be briefly summarized to give the reader some insight into this topic.

The studies reviewed have been divided into those dealing with environmental factors that might relate to underachievement such as family background, home conditions, social position of parents, community and neighborhood environment, companions and age preferences, counseling procedures, teacher influence, and social acceptance; and into those dealing with personal factors that might relate to underachievement such as aspirations, self-concept, emotional disturbance, health, intelligence quotient, goals and achievement motives, interests, values, and occupational choices.

Factors brought out in the literature on underachievement may be summarized as follows:

Environmental factors relating to underachievement that have been mentioned in the literature include:

1. Lack of good models in education and occupation in those they admire in home and community (44, 101).
2. Low educational level of family (44).
3. Unstable family structure (44).
4. Lower socioeconomic status (61, 40, 79, 90).
5. Ethnic and religious prejudice (44).
6. Rejection often by classmates for personal or caste reasons (6, 86).
7. Often over-age in their peer group, causing rejection (6).
8. Environment too complicated for their ability (73).
9. Have often received coercive influence from teachers or others (35).
10. Suffer more pressures (48).
11. Blocks to learning have been set up (48).

Personal factors related to low achievement and underachievement that have been mentioned in the literature include:

1. Underachievers dislike or fear people – especially parental or authoritative figures, siblings, and peers (44, 120).
2. Unable to create warm relationships with others (20, 16).
3. Have lower or unrealistic levels of aspiration (120, 104).
4. More often follow vocational goals set for them by others (2).
5. Have fewer vocational and other interests (2, 73).
6. Have stated goals not in line with their dominant interests (2, 73).
7. More prone to develop mental illness; more emotionally disturbed; have more difficulty in making non-academic adjustments (64, 109, 114, 16).
8. Have lower threshold of frustration (35).
9. Have lower intelligence quotient (64, 60).
10. Have low self-concept of ability (18, 44, 16).
11. Prefer to avoid conflicts (2).
12. Less cooperative, less dependable, have poorer judgment (2).
13. Prefer outdoor activity (2).
14. Prefer companions older than self (2).

CHAPTER 3

GENERAL ACHIEVEMENT

NOW THAT we have reviewed previous research relating achievement to ability grouping let us examine that portion of the Utah study that is concerned with this topic.

WORKING HYPOTHESES

The working hypotheses on achievement which formed the basis for designing this phase of the project were founded primarily upon psychological rationales developed by the author or postulated in the writings of other psychologists and educators.[3] Previous research was also used as a basis for developing working hypotheses, but because of the design limitations characteristic of most previous work, one finds this work a poor foundation upon which to build hypotheses. The major working hypotheses relative to achievement were:

1. Superior pupils will achieve more in ability grouped classrooms in which the curriculum is adjusted primarily through changes in learning rate than they will in random grouped classrooms in which enrichment is the primary means of curriculum adjustment. This hypothesis has considerable basis in the findings of previous research. A number of studies such as the work of Provust (90), West and Sievers (117), and Herr (59) for example, have found greater achievement gains for superior pupils in ability grouped classrooms than for superior pupils in hetergeneous classrooms.

Educators have also made certain assumptions about the ability grouping situation which would tend to lead one to expect superior pupils to make greater gains under this treatment. It is often maintained, for example, that the superior pupil is seldom challenged in the random grouped classroom and as a result, tends to work at a level far below his potential. In contrast, the ability grouped classroom is looked upon as providing a situation in which the superior pupil is faced with an accelerated program and more intense competition from his peers and these factors tend to motivate him to higher achievement.

2. With respect to the achievement of average pupils in ability grouped and random grouped classrooms, it was our working hypothesis that no significant differences would emerge. Both the findings of previous research and the rationale underlying ability grouping tend to lead to this hypothesis. A number of studies including Worlton (123), Provust (90) and Billet (9), have obtained greater differences in achievement for pupils of high or low ability than were found for pupils around the center of the ability distribution. A frequent criticism of heterogeneous grouping is that the curriculum is geared to the average pupil. If this is the case, the average pupils would have the least to gain in an ability grouped situation because the random grouped situation is already adapted to his level. Some educators have maintained that the average pupil suffers in the ability grouped situation because he is deprived of the assistance of superior pupils in the classroom situation. Some research, however, (74) suggests that little real contact takes place between pupils of different ability levels, thus minimizing the possible gains that could occur through interaction between pupils of different ability.

3. It was hypothesized that slow pupils would make greater achievement gains in the random grouped situation. This working hypothesis was somewhat harder to establish than those previously discussed, because of a number of apparently conflicting pressures and motivations that seem to be operating upon the slow pupil in the two grouping situations. Some of these pressures would appear to raise the achievement of the slow pupil in the ability grouped classroom, while others would appear to lower it. For example, the fact that the rate of learning in ability grouped classrooms can better be adjusted to the capacities of the slow learning pupils would lead one to expect these pupils to achieve more in the ability grouped situation. On the other hand, the stigma that most educators believe to be associated with membership in a slow group might well develop unfavorable attitudes, reduce motivation, and lower the self-concept and aspiration level of slow pupils in ability grouped classes.

In random grouped classrooms it has been argued that the class goals are so far beyond the ability of the slow pupil that he becomes discouraged and disinterested. On the other hand, it may be contended that the random grouped classroom in effect provides an accelerated situation for the slow pupil, the pressures of which might lead him to greater achievement, just as similar pressures in ability grouped classroom are thought to motivate the bright pupil. The reader may have noted in the review of the literature that studies dealing with the achievement of ability grouped slow pupils report somewhat conflicting results. Some studies have found no differences between the achievement of slow pupils in ability grouped and random grouped situations (85, 58, 90).

Others have shown relatively greater gains for slow pupils in ability grouped classes (9, 92). A study by Cook (27) found that ability grouped slow pupils achieved more in some subjects while random grouped slow pupils achieved more in others.

PROCEDURE

Sample

Achievement data were collected for Samples IV, VI, VII, VIII, and IX. Achievement gains from year to year as well as gains over the entire four years were studied for each sample. This procedure provided a large body of data on the nature of long term achievement in the two grouping treatments. Also, because achievement data for two or more samples were available at most grade levels, extensive cross checking of the results was possible.

Achievement Measures

Because of financial limitations imposed during the pilot study year, the California Achievement Test Battery already in use in the two school districts was employed. This battery, however, was not used beyond the pilot study year. A careful study of all available achievement tests was made in order to select the battery to be used during the final three years of this study. The Sequential Tests of Educational Progress (STEP) were finally selected. These tests have a number of advantages which were not possessed to the same degree by any of the other achievement batteries examined. Among these advantages are the following:

1. They emphasized broad understanding and the ability to utilize learned skills in solving new problems rather than ability to handle only specific facts, thus reducing the likelihood of content validity differences for the two districts participating in the research.
2. The battery contains vertically comparable tests that cover the entire grade span of subjects in this study. As most of the subjects in this research were administered two different levels of the STEP during the four years of the study, this vertical comparability is a highly desirable feature.
3. Detailed studies of the norm distributions carried out by Educational Testing Service indicated that in every instance, the test provided adequate ceiling for better students in the schools from which the norm samples were drawn and in general provided an adequate floor for low ability students. This characteristic is particularly important in a study in which pupils of a given grade level but of different ability levels must be analyzed separately.

ACHIEVEMENT RESULTS

Most of the differences in achievement gains will be reported between pupils of comparable ability in the two districts. The analyses were designed to yield information most pertinent to the question: "Do comparable pupils in ability grouped and random grouped classrooms make comparable achievement gains?" The usual analysis procedure was to identify pupils in District A and District R who were in a given ability level classification (superior, average, or slow) and carry out a separate analysis of covariance for each pair of comparable ability groups.

Sample IV

The first achievement battery was administered in October, 1958, when Sample IV pupils were beginning the fourth grade. This sample was studied from the beginning of the fourth grade to the end of the seventh grade and provides the best evidence that this project has to offer concerning the effects of ability grouping on the achievement of the elementary school child. The second battery was administered in April, 1959, and the results of this testing were adjusted for initial differences using analysis of covariance. The final adjusted overall achievement means were significantly higher for superior pupils in District A as compared to District R. This difference was also found in comparing average pupils. The difference between slow pupils in the two districts, however, was not significant, and in this case, the trend favored the random grouped slow pupils.

In addition to comparisons of overall achievement based on the total California Achievement Test Battery, pupils were compared on the basis of the achievement in the three subtests in the battery (see Table 4). Superior pupils in District A were significantly higher than comparable pupils in District R on all three of the subtest scores. Comparisons of the average pupils in the two districts showed District A pupils to be significantly higher in the language and reading subtests. Comparisons between slow pupils in the two districts on the three subtests of the CAT battery revealed no significant differences.

In comparing the first year's achievement of District A and District R, several points should be kept in mind. One important factor that probably influenced the scores in District A is the fact that these data were collected during the first year that ability grouping was employed in this district. Interviews with teachers and supervisors during this year indicated that teachers had not adjusted their techniques and materials to the different ability levels during this year as in later years. Another factor that almost surely operated to some degree in District A during the first year is Hawthorne Effect. The ability grouping system received considerable publicity and the local press favored this innovation. The change aroused interest and enthusiasm on the part

TABLE 4

ACHIEVEMENT OF SAMPLE IV PUPILS DURING THEIR FOURTH GRADE YEAR

	Arithmetic				Language				Reading			
Group	N	Initial Mean	Final Mean	$F_{y.x}$	N	Initial Mean	Final Mean	$F_{y.x}$	N	Initial Mean	Final Mean	$F_{y.x}$
Ability Grouped												
Superior (Dist. A)	80	4. 85[a]	620		80	5. 59	7. 04		81	4. 93	6. 94	
				13. 55				20. 17				50. 35
Random Grouped												
Superior (Dist. R)	155	4. 88	5. 91		160	5. 54	6. 53		160	4. 99	6. 29	
Ability Grouped												
Average (Dist. A)	86	4. 30	5. 39		83	4. 86	5. 87		86	3. 72	5. 67	
								7. 16				19. 74
Random Grouped				N. S.								
Average (Dist. R)	83	4. 35	5. 30		83	4. 51	5. 26		81	3. 88	5. 21	
Ability Grouped												
Slow (Dist. A)	62	3. 80	4. 68		59	3. 70	4. 16		60	3. 18	4. 27	
Random Grouped				N. S.				N. S.				N. S.
Slow (Dist. R)	46	3. 90	4. 76		46	3. 81	4. 22		46	3. 06	4. 38	

[a] CAT Grade Placement Scores

[b] All F values given are significant at the . 01 level.

In the following tables, F or t values significant at the . 05 level will be indicated *. Those not so marked are significant at the . 01 level.

of many teachers and administrators in District A. Such interest usually can be expected to bring about some favorable change in pupil performance regardless of the merits of the variable introduced.

Achievement gains for the second year of the study were computed by comparing the scores on the CAT battery administered near the end of the fourth grade with a battery made up of four tests from the Sequential Tests of Educational Progress, (STEP) obtained near the end of the fifth grade. The STEP and CAT have two subtest areas in common, mathematics and reading. Gains in these two areas and on the overall battery scores were compared for this sample. Superior pupils did not differ significantly in mathematics gains or in overall gains based upon the composite battery scores. The superior pupils in District R, however, made significantly greater gains in reading (t = 2. 29) than comparable pupils in District A. Comparisons of average subsamples revealed that District A pupils made significantly greater gains (. 01 level) in reading, mathematics and overall achievement. Slow pupils in District R registered significantly greater gains in reading (t = 3. 65) and in the overall battery score (t = 4. 44). The difference in mathematics gains, though not statistically significant, also favored the District R sample. These differences along with other achievement differences between the two treatments are summarized briefly in Table 6.

During the third year of the study, both the initial and final achievement means were based on the STEP battery. It was possible during this year to analyze achievement gains on the overall battery score and also on the mathematics, science, reading, and social studies subtests. The gains made by pupils of similar ability in District A as compared with District R revealed no significant differences in mathematics, reading, social studies, or in the overall STEP battery score. Only in science did statistically significant differences occur. These differences were present at all three ability levels and in every case favored District A.[4]

During the fourth year of the study, pupils in Sample IV were in the seventh grade. It will be recalled that at the seventh grade level, District A employed ability grouping only in mathematics and science. The STEP battery from 3B was administered near the end of the seventh grade. Superior pupils in both districts were given the high school form of the STEP (2B) in order that there would be no chance of their being restricted by the test ceiling. The results revealed only one significant difference in comparisons between the ability grouped and random grouped seventh grade pupils of similar ability levels on the STEP mathematics and science tests. It was found that superior ability grouped pupils made significantly greater gains in mathematics than superior random grouped pupils (t = 2. 21). The District R slow pupils made greater achievement gains in both tests than the District A slow pupils although these differences were not large enough to be statistically significant. For average pupils the differences between the two districts were inconsistent and small.

Results for the Four Years

In addition to analyzing gains year by year for Sample IV, the scores on the initial CAT battery administered in the fall of 1958 were compared with the scores made by the same pupils on the final STEP battery administered in spring of 1962 (see Table 4). These results indicated that over the four year period, the ability grouped superior sample had made significantly greater overall achievement gains. For the average groups, there were no significant differences, and for the slow groups, the District R pupils made significantly greater gains. For this comparison the pupils were classified in accordance with their ability level at the outset of the study while in the year to year comparisons, classification was based upon the pupil's ability level at the beginning of the year in question.

The year by year analyses appear to have one disadvantage. They fail to reflect adequately the cumulative effects of the two grouping treatments because pupils shift in their ability level classification from year to year if their performance justifies a change. For example, a child who is initially in the slow group in District A may be moved to the average group for the second year of the study and will thus become part of the average group during the second year's analysis. There was, in fact, a rather large number of pupils in District A who did move out of the slow group. The data for the entire four years fill in this gap, and provide a more accurate picture of long term gains under the two grouping treatments.

Sample VI

This sample was tested at the beginning of the sixth grade in October, 1958 and was followed through the ninth grade with final achievement measures administered in April, 1962. Sixth grade achievement for this sample was measured by the California Achievement Test Battery. The same limitations apply to these data that were discussed previously with regard to Sample IV. It may be seen in Table V that the initial mean scores of District A and District R pupils at the three ability levels were closely comparable. Analysis of variance of the total battery scores yielded F values of . 24, . 67, and . 52 between slow, average, and superior pupils in the two districts. The final overall achievement means were adjusted for these slight initial differences using covariance analysis. Adjusted final means showed that superior pupils in District A had made considerably greater gains (. 01 level) on the CAT composite battery score than their peers in District R. Differences significant beyond the . 01 level were

TABLE 5

ACHIEVEMENT OF SAMPLE VI PUPILS DURING THEIR SIXTH GRADE YEAR

Group	Arithmetic				Language				Reading				Total Battery			
	N	Initial Mean[a]	Final Mean[a]	$F_{y.x}$	N	Initial Mean	Final Mean	$F_{y.x}$	N	Initial Mean	Final Mean	$F_{y.x}$	N	Initial Mean	Final Mean	$F_{y.x}$
Ability Grouped																
Superior (Dist. A)	125	6. 55	9. 52		125	7. 10	8. 55		125	7. 01	8. 77		141	6. 77	8. 96	
Random Grouped				90. 82				14. 62				103. 64				41. 03
Superior (Dist. R)	194	6. 40	8. 48		197	7. 08	8. 20		196	7. 21	8. 02		244	6. 90	8. 23	
Ability Grouped																
Average (Dist. A)	114	5. 83	7. 62		117	6. 29	7. 37		116	5. 81	7. 65		136	5. 79	7. 52	
Random Grouped				5. 10				N. S.				29. 95				20. 4
Average (Dist. R)	119	5. 82	7. 33		125	6. 08	7. 05		124	5. 63	7. 04		153	5. 84	7. 14	
Ability Grouped																
Slow (Dist. A)	110	5. 11	6. 40		108	5. 28	6. 07		109	4. 66	6. 07		128	4. 95	6. 13	
Random Grouped				N. S.				13. 50				N. S.				N. S.
Slow (Dist. R)	69	5. 15	6. 28		65	5. 24	5. 60		65	4. 57	5. 88		88	4. 92	5. 97	

[a] CAT Grade Placement

TABLE 6

SUMMARY OF ACHIEVEMENT DIFFERENCES BETWEEN DISTRICT A AND DISTRICT R SAMPLES OVER THE FOUR YEARS OF THE STUDY

	Sample IV						Sample VI				Sample VII				Sample VIII				Sample IX			
	1st Yr.	2nd Yr.	3rd Yr.	4th Yr.	4 Yrs.	Last 3 Yr.	1st Yr.	2nd Yr.	3, 4 Yrs.	4 Yrs.	1st Yr.	2nd Yr.	3, 4 Yrs.	4 Yrs.	1st Yr.	2nd Yr.	3, 4 Yrs.	4 Yrs.	1st Yr.	2nd Yr.	3, 4 Yrs.	4 Yrs.
Superior																						
Math.	A[a]	NS	NS	A			A	R	NS	A	NS	A	NS	NS	NS	A	NS	NS	NS	NS	A	NS
Lang.	A						A															
Science			A	NS				NS	NS	NS		A	NS	NS		NS	NS	NS		NS	NS	NS
Reading	A	R	NS				A															
Soc. St.			NS																			
Overall	A	NS	NS		A	NS	A															
Average																						
Math.	NS	R	NS	NS			A	NS	NS	NS	R	NS	NS	NS	R	NS	NS	NS	NS	NS	A	A
Lang.	A						NS															
Science			A	NS				NS	A	A		A	NS	NS		A	NS	NS		NS	NS	NS
Reading	A	R	NS				A															
Soc. St.			NS																			
Overall	A	R	NS		NS	R	A															
Slow																						
Math.	NS	NS	NS	NS			NS	R	NS	NS	R	NS	NS	NS	NS	A	NS	NS	NS	NS	NS	A
Lang.	NS						A															
Science			A	NS				NS	NS	NS		NS	R	NS		NS	NS	NS		NS	NS	NS
Reading	NS	R	NS				NS															
Soc. St.			NS																			
Overall	NS	R	NS		R	R	NS															

[a] A – District A sample made significantly greater achievement gains.
R – District R sample made significantly greater achievement gains.
NS – Difference between Districts A and R not significant.

also found between average pupils in the two treatments, again favoring District A. No significant differences were found between slow pupils. The pattern of differences for this sample is very similar to that found in the first year performance of Sample IV (see Tables 4 and 5).

Achievement on the three specific subtests of the CAT: arithmetic, language, and reading were also compared. These results also followed a pattern generally similar to that found in the Sample IV analysis. District A superior pupils had significantly higher final adjusted mean scores on all three of the CAT subtests. These differences were all significant beyond the .01 level with the arithmetic and reading differences amounting to over one-half year in terms of grade placement scores and yielding very large adjusted F's (90.82 and 103.64 respectively) on the analysis of variance. It will be noted that the greatest difference between the two superior groups was in reading as was also the case for Sample IV. A comparison of average pupils in the two districts revealed significant differences favoring District A pupils in arithmetic and reading. These differences were much smaller than those found for superior pupils. No significant differences were found between the average groups in the language subtest. District A slow pupils were significantly higher in their final adjusted language means, but there were no significant differences between slow pupils in the two districts on the arithmetic or reading subtests.

In order to determine achievement gains during the seventh grade, the CAT scores earned by Sample VI pupils at the end of the sixth grade were used to adjust the means on the STEP mathematics and science tests that were administered at the end of the seventh grade.

The results showed the District R superior and slow pupils to be significantly higher than comparable District A pupils in mathematics at the end of grade seven ($F_{y.x}$ for superior 4.01, for slow 11.69). This difference, though not large for the superior sample, is noteworthy as it cancels out the rather large difference favoring the District A pupils that was found during the first year of the study. The STEP mathematics scores earned by the two samples at the end of grade seven were less than one point apart, indicating that, when one considers the first two years for this sample, achievement in mathematics was nearly identical for superior pupils in the two districts. In contrast, the significant final mean differences between the slow pupils leave the District R sample somewhat ahead in mathematics achievement for the initial two years.

Because the CAT does not include a science score, STEP science scores were adjusted for initial differences using the CAT composite battery score. No significant differences in the adjusted final STEP science mean scores were found between any of the comparable ability groups in the two districts.

No achievement measures were administered to the secondary school sample during the third year of the study, so the achievement during the final two years of the study is covered in one analysis. Gains made between the end of grade seven and the end of grade nine on the STEP achievement battery were compared using analysis of covariance. No significant differences were found between comparable ability groups in mathematics. A small difference favoring the District R average sample was found in analysis of the science gains ($t = 2.26$). Gains at the superior and slow ability levels were not significantly different for the two grouping systems.

Results for the Four Years

As three of the four years covered in this analysis involved secondary school achievement, analysis of gains in mathematics and science seemed more appropriate than an analysis of overall gains such as was used in the four-year analysis of Sample IV achievement. In mathematics, the superior District A sample showed a small but significantly greater gain over the four-year period ($F_{y.x} = 5.82$). This difference takes on greater importance when we recall that District A superior pupils in this sample were significantly lower in socioeconomic status and thus probably less motivated to achieve than the superior District R pupils. The District A average pupils registered a similar gain in science over their District R peers ($F_{y.x} = 4.46$). All other differences failed to reach statistical significance. The Sample VI data suggest that achievement gains that can be attributed to the ability grouping system over four years are generally small.

Sample VII

This sample was tested at the beginning of the seventh grade in October, 1958, and was followed through the tenth grade with final achievement measures administered in April, 1962. Along with Sample VI, these pupils represent the best data provided in this research concerning the effects of ability grouping at the junior high school level. During three of the four years of this study, pupils in these two samples were at the junior high school level.

The first year analysis for this sample was limited to mathematics as no science measure is included in the CAT battery. Comparisons of superior pupils based on analysis of covariance yielded no significant differences. District R average pupils showed significantly higher final means on the CAT arithmetic tests than comparable District A pupils ($F_{y.x} = 9.50$).

Achievement during the eighth grade for Sample VII was determined by adjusting end-of-year STEP achievement means in mathematics and science for initial differences found in the California Achievement Test Battery administered at the end of the seventh grade. Comparisons between final adjusted mean scores in mathematics showed a significant difference in favor of the District A superior pupils ($F_{y.x} = 6.18$). Comparisons between average and slow pupils in mathematics achievement yielded no significant differences. The final adjusted mean scores in science also showed statistically significant differences in favor of the District A superior pupils ($F_{y.x} = 6.39$). District A average pupils were also significantly higher ($F_{y.x} = 6.04$). No significant differences were found between slow pupils in the two districts.

Achievement of the Sample VII pupils between the end of the eighth grade and the end of grade ten probably provides the most reliable comparative data for this sample. These data include adequate numbers of cases, extend over a two-year period and compare gains on two different forms of the *STEP* mathematics and science tests. Comparisons in mathematics and science yielded only one significant difference between pupils of comparable ability in the two districts. In science achievement the District R slow pupils were significantly higher than comparable District A pupils ($F_{y.x} = 7.21$).

The overall gains in mathematics and science achievement for the four years of the study were also computed for Sample VII. No significant achievement differences were found between District A and District R at any of the three ability levels in either science or mathematics. The only cumulative four year difference that approached statistical significance was between the final adjusted science means of superior pupils, and favored the District A sample.

Sample VIII

This sample was tested at the beginning of the eighth grade in October, 1958, and was followed through the eleventh grade with final achievement measures administered in April, 1962. As was the case with Sample VII, the first year's achievement analysis for Sample VIII was limited to mathematics achievement. End-of-year achievement based on the *CAT* arithmetic test was adjusted for initial achievement differences on a different form of the same test.

Comparisons of the final adjusted means for the first year indicated a significant difference in favor of District R average pupils ($F_{y.x} = 4.62$). Differences between District A and District R superior and slow pupils were not statistically significant. During the second year of the study, Sample VII pupils were in the ninth grade. District A superior and slow pupils obtained significantly higher adjusted final mean scores in mathematics than comparable District R pupils ($F_{y.x} = 4.35$ and 5.52). No significant differences were found between average pupils in the two districts in mathematics achievement. In comparing science achievement, only one significant difference was found; District A average pupils making greater gains ($F_{y.x} = 10.54$).

The achievement made by Sample VIII pupils between the end of the ninth grade and the end of the eleventh grade in mathematics and science was measured by two forms of the *STEP* achievement test in these areas. Because of the wide range of electives available during the high-school years, these data can be considered only tentative evidence on the relative merits of the two grouping treatments. No significant differences were found between the final adjusted means of District A and District R pupils in either mathematics or science at any of the three ability levels.

Results for the Four Years

A comparison of achievement of Sample VIII pupils over the entire four years of the study was made using analysis of covariance. None of the differences in final adjusted means in mathematics or science between comparable District A and District R pupils were statistically significant. When it is remembered that these comparisons represented the cumulative effects of four years under the ability grouping and random grouping treatments, the lack of any differences reaching the .05 level certainly suggests that the two systems had little or no differential effect upon the long term achievement gains of these pupils.

Sample IX

This sample was tested at the beginning of the ninth grade in October, 1958, and was followed through the end of the twelfth grade when final achievement measures were administered This group was somewhat smaller than samples at the other grade levels, but represents the best evidence available in this research concerning the relative achievement gains of ability grouped and random grouped pupils in mathematics and science at the high school level. During three of the four years of the study, these pupils were high school students. Limitations imposed by the range of mathematics and science electives at the high school level should be considered when interpreting Sample IX achievement results.

Achievement data during the ninth grade was limited to mathematics achievement on the *California Achievement Test*. No significant differences were found between the ability grouped and random grouped pupils during this year. Analysis for the second year of the study was based on the *CAT* battery administered at the end of the ninth grade and the *STEP* mathematics and science tests administered at the end of the tenth grade.

No significant differences were found between District A and District R pupils of comparable ability in either mathematics or science achievement. Achievement during the period between the end of the tenth and the end of the twelfth grade was measured by two forms of the STEP mathematics and science tests. In mathematics, the District A superior group had a significantly higher final adjusted mean score ($F_{y.x} = 5.51$). Ability grouped pupils at the average level also had a significantly higher adjusted final mean score in mathematics ($F_{y.x} = 6.57$). No significant difference was found between slow pupils in the two grouping systems. Comparisons of science achievement for District A and District R pupils of similar ability revealed no significant differences over this two year period.

Results for the Four Years

Analysis of the achievement gains for the four years of the study revealed significant differences in final adjusted mathematics achievement means, favoring District A average and District A slow pupils as compared with District R pupils of similar ability ($F_{y.x} = 6.14$ and 77.8). No significant differences were found in mathematics between superior pupils in the two districts. Comparisons between treatments on final adjusted science achievement yielded no significant differences.

SUMMARY AND CONCLUSIONS

In the preceding pages we have examined in some detail the achievement data collected on the five samples of pupils studied in this research during the four years of the project. Now let us attempt to pull these data together in order to appraise the effect of ability grouping and random grouping upon achievement in the three major school divisions – elementary, junior high, and senior high school.

Elementary School Achievement

Elementary school achievement data in this research are based on the performance of Sample IV in grades four, five, and six; and Sample VI in grade six. Achievement in four subject areas as well as overall achievement were compared during the elementary school years for these samples. The reader will note in Table 6 that during the elementary school years, a total of 54 statistical comparisons were made between District A and District R pupils. Of these 54 comparisons, 28 were statistically significant. Nineteen of these significant differences were found favorable to District A pupils while nine favored District R pupils. At first glance this would seem to indicate a rather marked overall advantage for ability grouping at the elementary level. On looking further, however, we find that of the 19 significant differences favoring District A pupupils, 15 occurred during the first year of the study. It is interesting to compare the sixth grade results obtained for Sample IV and for Sample VI and speculate about the reason for the much larger differences found among Sample VI pupils. There are several possible reasons for these differences, but the most likely appears to be that Hawthorne Effect was operating rather strongly in favor of the District A program during the first year. If these first year overall differences had been due primarily to a true superiority of ability grouping over random grouping we would expect the differences to grow larger each year as the cumulative effects of the better system widened the achievement gap. It will be seen in Table 6 that no such cumulative effects occurred and in fact, most of the achievement differences favoring District A disappeared by the time Sample IV pupils had completed the sixth grade.

Although the data obtained in this study show no great overall superiority for either ability grouping or random grouping, some interesting differences in the two grouping treatments emerge when we consider the data for each ability level separately. When we compare superior pupils in the two districts, we find that for the 18 achievement comparisons, 11 are statistically significant with ten of these favoring the District A pupils. In terms of overall achievement differences for the four years of the study, District A superior pupils were significantly higher.

Comparisons between average pupils during the elementary school years showed no consistent trend favoring either grouping system. Of the 18 comparisons made, seven were significant favoring District A, four favoring District R and the remainder were not significant. Thus, there appears to be little to choose between random grouping and ability grouping so far as the achievement of average pupils is concerned.

Comparisons between slow pupils in the two grouping treatments showed six significant differences, four of these favoring District R and two favoring District A. District R slow pupils made significantly greater overall achievement gains during the four years of the study than their peers in District A. If, as seems likely, Hawthorne Effect were operating strongly in favor of District A during the first year, the relative performance of District R slow pupils was probably even better than indicated in Table 6.

All in all, we may conclude that neither ability grouping with acceleration nor random grouping with enrichment is superior for all ability levels of elementary school pupils. In general, the relative achievement advantages of the two grouping systems were slight, but tended to favor ability grouping for superior pupils and random grouping for slow pupils. As was hypothesized, the differences for average pupils did not consistently favor either grouping treatment.

Junior High School Achievement

All five of the samples employed in this study were in junior high school at some time during the four years of the project. When we combine the achievement data for these five samples, we find that a total of 60 statistical comparisons were made between comparable groups in the two districts; 33 in mathematics and 27 in science. Of the mathematics comparisons, five were significant in favor of District A subgroups, five in favor of District R, and the remaining 23 were nonsignificant. Of the science comparisons, five were significant favoring District A; one favoring District R and the remaining 21 were nonsignificant. Thus, in terms of the overall effects of the two grouping treatments, the districts appear comparable in mathematics achievement, but there is some tendency for District A pupils to make greater achievement in science.

Let us now study the results of the junior high data at each of the three ability levels. A total of 11 comparisons were made between superior District A and R students during the junior high school years. Of these, four were significant in favor of District A, while one favored District R. Of the nine comparisons between science achievement of superior pupils, only one difference was statistically significant, favoring District A.

For average pupils, it will be noted in Table 6 that two significant differences favored District R in mathematics, and the remaining nine comparisons are nonsignificant. The science achievement of average pupils shows somewhat better achievement for District A pupils with four of the nine comparisons being statistically significant, all favoring District A.

In comparing the mathematics achievement of slow pupils, we find one significant difference favoring District A, and two favoring District R. In science, only one of the nine comparisons between slow groups was statistically significant, this favoring District R. From the aforementioned data we may conclude that at the junior high school level, there is some tendency for ability grouping to lead to greater mathematics achievement among superior pupils and greater science achievement among average pupils. Among slow pupils, the differences between the two grouping treatments are small, but tend to favor District R in both mathematics and science.

Senior High School Achievement

Somewhat less data are available in the Utah study on senior high school achievement. These data are based on the tenth and eleventh grade achievement of Sample VIII, and the tenth, eleventh and twelfth grade achievement of Sample IX. A total of thirty comparisons were made between the mathematics and science achievement of District A and R pupils in these samples. Of these thirty comparisons, four were statistically significant favoring District A, and the remaining 26 were nonsignificant. All four of the significant differences were in mathematics achievement with one difference between superior pupils, two between average pupils and one between slow pupils in the two districts.

These data would lead to the tentative conclusion that ability grouping is associated with slightly greater mathematics achievement at all ability levels during the high school years. It may also be concluded that the grouping treatments have no effect upon science achievement during these years. In reviewing the high school achievement data, the reader should recall that these data can be accepted with less confidence than either the elementary or junior high school data for the reasons previously discussed.

CHAPTER 4

OVERACHIEVERS AND UNDERACHIEVERS[5]

ALTHOUGH a number of studies have been conducted in which achievement under different grouping systems has been compared, achievement of individual pupils in these systems as related to the achievement one might expect on the basis of the individual's ability has not been explored. In this phase of the Utah study, samples of elementary school pupils from random grouped and ability grouped classrooms were classified as overachievers, underachievers and normal achievers based on the relationship between their achievement, chronological age, and mental age. The proportion of overachievers, normal achievers and underachievers among pupils of different ability levels in the two systems were then studied in an effort to obtain insight into the question of who is motivated to achieve under these two methods of grouping.

WORKING HYPOTHESES

The following working hypotheses were established with respect to the performance of overachievers, underachievers, and normal achievers in ability grouped and random grouped classrooms:

1. There will be no significant difference in the overall proportion of overachievers, underachievers, and normal achievers in the ability grouped sample as compared with the random grouped sample.
2. There will be no significant differences in the proportions of overachievers, normal achievers and underachievers between the fourth and sixth grade samples.
3. There will be significantly more boys than girls who are classified as underachievers and fewer boys who are classified as overachievers.
4. There will be a significant tendency for superior pupils in the ability grouping system to be more frequently overachievers and less frequently underachievers than in the random grouping system. This hypothesis is based on the observation that schools employing random grouping often slant the level of work to the average pupil and in so doing may fail to challenge the superior pupil.
5. There will be no significant difference between the ability grouping system and the random grouping system in the proportion of overachievers, normal achievers, and underachievers among average pupils. Because of the tendency to aim the curriculum at the average pupil in the heterogeneous class, the two grouping treatments involved in the Utah study would be most similar for pupils at this ability level. One difference between the treatments that may affect the average pupil, is his association with superior pupils in the heterogeneous class. Some educators have maintained that the average pupil is motivated to greater achievement because of the presence of superior pupils in the heterogeneous classroom. If this claim is valid, more overachievers and fewer underachievers should be found in the District R average sample.
6. There will be significantly fewer underachievers and more overachievers among slow pupils in the random grouping system than in the ability grouping system. This hypothesis is based on the same rationale as hypothesis four. It is anticipated that random grouping, by slanting instruction to the average pupil, will tend to put pressure on the slow pupil to achieve.

PROCEDURE

Sample

The data collected in the Utah study that deal with achievement potential are based upon measures obtained on pupils in Samples IV and VI during the first year of the study. A total of 674 District A pupils and 793 District R pupils were compared. These pupils were classified in terms of achievement category (overachievers, normal achievers, and underachievers) sex, and ability level (superior, average, and slow).

Measures

There is no general agreement among educators on a precise definition of "overachiever" and "underachiever". Most definitions, however, provide for comparisons between the individual's achievement and his achievement potential as estimated by an intelligence test. Underachievers are defined as achieving at some point below their potential achievement. Overachievers are less frequently discussed. Most educators have difficulty accepting the concept of the "overachiever" as this implies that the individual is achieving more than his potential. In this study we have defined the underachiever as a person who is achieving more than nine months below the normal level expected on the basis of mental age and chronological age, and the overachiever as a person achieving the same amount above this expected level. The expected level, however, should not be thought of as the individual's maximum potential. The concept of the overachiever becomes more acceptable if we consider the following analogy. Let us say that the normal achiever is similar to the average automobile that has eight cylinders but generally operates on only five or six of them. The underachiever then becomes comparable to the auto that lurches unsteadily along on three or four of its eight cylinders while the overachiever may be compared with that rare vehicle that speeds along on seven or perhaps even all eight cylinders. Thus, the overachiever performs better than the average achiever of comparable ability, and more closely approaches his achievement potential.

In order to identify underachievers, the mean achievement score of fourth and sixth grade pupils on form W of the California Achievement Test Battery administered at the end of the school year was calculated for both districts. The overall mean grade placement score on the CAT for the District A sixth grade group was 7. 58 as compared with the overall mean grade placement for the District R group of 7. 48. Comparable means for the fourth grade were 5. 67 and 5. 68. The classification of pupils as underachievers and overachievers presents some problems. A number of methods have been used from time to time including the rating of any pupils who are not achieving up to their normal grade placement as underachievers. Such a system ignores the pupil's achievement relative to his ability. Another system defines the underachiever as a pupil whose educational age is lower than his mental age. This system seems to place excessive weight on ability and tends to classify the majority of bright pupils as underachievers. Although the bright pupil may be expected to score well above his grade placement in subjects such as reading that are not closely tied to the learning of specific sequences of operations, it is not reasonable to expect the bright pupil to score greatly above his grade placement in subjects such as mathematics where a knowledge of certain operations is essential for working the more advanced problems.

In the Utah study a system of classification for underachievers developed by the Santa Monica California Unified School District was employed. The formula used in this case determines the expected grade placement by taking the sum of twice the pupil's mental age plus his chronological age and dividing by three. The mental age used in this study was based upon the Language Score of the California Test of Mental Maturity (CMM). This formula is, of course, arbitrary, but seems to be reasonable as it takes into account both mental age and chronological age in arriving at an expected educational age score. Each pupil in the sample was given a score that represented the difference between his expected educational age in months, based on the Santa Monica formula and his obtained educational age in months based on the California Achievement Test Battery, form W. This battery was found to yield educational age scores consistently higher than expected. The median educational age on the CAT was four months higher than the median expected educational age calculated by the Santa Monica formula. The median educational age was used as the center of the normal achieving group. Pupils whose educational age was nine months or more above this median were classified as "overachievers", while those whose educational age was nine months or more below this median were classified as "underachievers".

RESULTS

The significance of differences in the proportions of overachievers, normal achievers, and underachievers in the two grouping treatments was determined using chi square contingency tables. Overall comparisons between pupils in the two grouping treatments were made as well as comparisons between the sexes, between pupils in the fourth versus sixth grades, and between District A and District R pupils at a given ability level.

Sex Differences

The first comparisons that were made were aimed at determining whether sex differences occurred in the overall proportion of overachievers and underachievers in the two treatments. Chi square contingency tables were set up comparing boys with girls in the proportion of underachievers, normal achievers, and overachievers. In all ten of the comparisons made, there was a greater proportion of overachievers and lesser proportion of underachievers among girls. Seven of these ten differences were statistically significant (see Table 7).

Grade Differences

The next step in the analysis involved determining whether differences existed between the proportion of overachievers, normal achievers and underachievers in the fourth grade as contrasted with the sixth grade samples within each grouping treatment. An examination of the chi square contingency tables revealed that in every case there was a trend toward fewer overachievers and a greater proportion of underachievers from the fourth to the sixth grades. In District A, this tendency was significant at the .01 level for both boys and girls. In District R, the same trend existed although it was somewhat more marked in the case of the girls. This trend toward a drop in the proportion of pupils who were achieving up to or beyond expectation and an increase in the number of pupils who were achieving below expectation may have been the result of a change in the pupils' value system that occurred between these two grade levels. At the fourth-grade level, most pupils are still oriented strongly toward the teacher and are likely to value school achievement more highly because of the teacher's influence. By sixth grade, however, the pupil's orientation has shifted to a considerable extent to the peer group in which school achievement may be valued very little or may even be ridiculed as a sign of submission to adult authority.

Between-Treatment Differences

Although the sex differences and grade differences we have discussed provide useful insights into the nature of underachievement, the main purpose of the study was to compare the proportion of overachievers, normal achievers and underachievers found in the two grouping treatments.

A noteworthy finding in comparing underachievers in the two systems of education is the markedly greater percentage of underachievers in the random grouped system. Of the 793 pupils in District R, 175 or 22 percent were classified as underachievers. In contrast, only 105 of the 674 District A cases or 15.6 percent were classified as underachievers. This difference is highly significant. As the courses of study and overall achievement levels within the two districts were very similar, the comparatively large number of underachievers in the random grouped sample may be an important factor in evaluating the relative merits of random versus ability grouping. Comparisons of the number of overachievers show even greater differences. The ability grouped sample had 29.1 percent overachievers as compared with 18.9 percent in the random grouped sample. This difference is highly significant beyond the .001 level.

The proportions of overachievers and underachievers among ability grouped and random grouped subsamples of comparable ability also showed some interesting differences (see Table 8). Examination of comparable subsamples indicated that the percentages of boys and girls in the districts were generally similar. Therefore, for this phase of the analysis, the sexes have been combined within each subsample. Fourth grade and sixth grade groups have been analyzed separately so that the results for one grade could be checked against the results for the other grade, and so that trends could be studied. To compare the proportions of overachievers and underachievers in comparable ability subgroups of the two districts, chi square contingency tables were set up in which superior, average and slow students in the ability grouped sample were compared with the same subgroups in the random grouped sample.

The comparison between superior pupils showed that the ability grouped sample had a greater proportion of overachievers and a smaller proportion of underachievers at both fourth and sixth grades. In the fourth grade sample, 47.2 percent of the ability grouped pupils were classified as overachievers as compared with 31.6 percent of random grouped pupils. Greater differences in the same direction were found in comparing the superior sixth grade pupils in the ability grouped and random grouped samples. Among ability grouped superior pupils, 41.3 percent classified as overachievers as contrasted with 13 percent of the superior pupils in random grouped classes. Differences between the two systems in the proportions of sixth grade superior students classified as underachievers were smaller with 16.7 percent of District A group and 20 percent of the District R group so classified.

The proportions of overachievers, normal achievers, and underachievers among average pupils in the two districts were not significantly different for either the fourth or sixth grade samples. A larger proportion of underachievers was found among the average pupils in random grouped classrooms but this difference was not large enough to establish a significant overall trend.

A comparison of slow fourth grade pupils in ability grouped and random grouped classes also showed no significant difference in the proportion of overachievers, normal achievers, and underachievers. These results were supported by data on slow sixth-grade pupils in the two samples which also failed to reveal significant differences.

SUMMARY AND CONCLUSIONS

On the basis of these findings, the following conclusions can be drawn:

1. There was a significant tendency for girls more frequently to be overachievers and less frequently to be underachievers than boys. This trend was present at both grade levels and in both treatments.
2. A comparison of fourth and sixth grade pupils in each of the grouping treatments indicated a significant trend for both boys and girls to achieve less at the sixth grade level relative to their potential. In all cases, the sixth grade samples had a greater number of underachievers and a smaller number of overachievers than comparable fourth grade samples. This drop was considerably greater for girls in District R than for girls in District A which might suggest that ability grouping tends to lessen the trend toward lower achievement between the fourth and sixth grades. This implication is not borne out by the results of the boys, however, whose drop in relative achievement was substantially the same in both the random grouped and the ability grouped samples.
3. Overall comparisons between Districts A and R indicated that ability grouped pupils were more often classified as overachievers and less often classified as underachievers than were random grouped pupils.
4. In comparing pupils at the three ability levels in the two districts, the ability grouped samples were generally found to have a more favorable dis-

TABLE 7

DIFFERENCES BETWEEN BOYS AND GIRLS IN THE DISTRIBUTION OF OVERACHIEVERS, NORMAL ACHIEVERS AND UNDERACHIEVERS

	Favored Group[a]	Chi Square	Significance Level
District A			
Superior Boys vs. Superior Girls	G	7.58	.05
Average Boys vs. Average Girls	G	3.87	NS
Slow Boys vs. Slow Girls	G	3.56	NS
Boys vs. Girls (overall) (4th grade)	G	8.45	.05
Boys vs. Girls (overall) (6th grade)	G	16.07	.001
District R			
Superior Boys vs. Superior Girls	G	8.63	.05
Average Boys vs. Average Girls	G	22.23	.01
Slow Boys vs. Slow Girls	G	3.00	NS
Boys vs. Girls (overall) (4th grade)	G	19.04	.001
Boys vs. Girls (overall) (6th grade)	G	17.36	.001

[a] "G" indicates girls had a more favorable distribution of overachievers, normal achievers and underachievers.

TABLE 8

DIFFERENCES BETWEEN DISTRICT A AND DISTRICT R PUPILS IN THE DISTRIBUTION OF OVERACHIEVERS, NORMAL ACHIEVERS AND UNDERACHIEVERS

	Favored Group[a]	Chi Square	Significance Level
4th Grade Superior Pupils	A	8.43	.05
6th Grade Superior Pupils	A	41.14	.001
4th Grade Average Pupils	A	3.93	NS
6th Grade Average Pupils	A	1.76	NS
4th Grade Slow Pupils	A	3.94	NS
6th Grade Slow Pupils	A	4.68	NS

[a] Indicates ability grouped pupils had a more favorable distribution of overachievers, normal achievers and underachievers.

tribution of overachievers and underachievers at both grade levels. The differences between superior pupils in the two districts were statistically significant at both grade levels with District A having more overachievers than District R and a smaller percentage of underachievers. These results seem to support the often stated contention that ability grouping presents a greater challenge to superior students.

5. No significant differences were found between the proportions of overachievers, underachievers, and normal achievers among pupils of average ability in the two districts. These results support the working hypothesis and suggest, as might be expected, that the two systems of grouping have little effect on the achievement or motivation of average pupils. In random grouping systems the curriculum is generally aimed at the average pupil. Thus, one of the oft-stated advantages of the ability grouping, that it fits the curriculum more accurately to the needs of the pupil, does not apply to the average groups.

6. Comparisons of the proportions of overachievers, underachievers and normal achievers among slow pupils in the two systems revealed no significant differences. For slow pupils, the ability grouping system is generally thought to have the advantage of presenting the student with more suitable curricular materials while the random grouping system might be expected to produce greater achievement because of the pressure put on the slow pupil to keep up with the rest of the class. If these factors are operating, it must be concluded that neither is of sufficient importance to lead to significantly different achievement patterns for the slow pupil.

The reader is reminded that the above conclusions are based upon achievement data collected during the pilot study year of the Utah project and should therefore be considered tentative until supported by additional research evidence.

CHAPTER 5

STUDY METHODS AND ATTITUDES

ALTHOUGH no previous research was found that was concerned directly with the effects of ability grouping and random grouping upon study habits, this area seemed a worthwhile one to explore.

WORKING HYPOTHESES

1. It was hypothesized that superior pupils in ability grouped classrooms would obtain significantly higher scores on the study habit measures than would superior pupils in random grouped classrooms. Many educators are convinced that ability grouping tends to develop better study habits. This belief is particularly strong with regard to superior pupils and is based on the idea that the superior pupil in a random grouped class is not sufficiently challenged and is capable of doing the assigned work with very little real effort. Thus, he tends to develop poor study habits and attitudes. Conversely, in an ability grouped classroom the bright student is forced to compete with other students of equal ability in a more challenging program, requiring him to develop better study habits in order to maintain his position in the class.

In view of the lack of evidence in this area and the absence of factors in the grouping treatments that would suggest a clear cut superiority for either treatment, there is little reason to expect that ability grouping would tend to develop better study habits among average or slow pupils. In fact, it might be argued that the average pupil would be motivated to study harder in the random grouped class where he is challenged by superior pupils. With regard to the slow pupil, the same argument can be made, but a counter argument is also possible to the effect that the slow pupil in the random grouped class is faced with competition he cannot hope to meet and may thus react with some form of rationalization that would tend to reduce his efforts rather than increase them.

2. It was hypothesized that there would be no significant differences between the scores of average and slow pupils in the two grouping treatments on the study habits measures.

PROCEDURE

Considerable evidence concerning study habits was collected over the four years of the Utah study. Pupils in Samples IV, VI, VIII, and IX were administered study habits measures at grade levels ranging from seven to 12. Two measures were used, the California Study Methods Survey (CSMS) and the Survey of Study Habits and Attitudes (SSHA). The CSMS contains 150 items and yields three subscores and total score. The subscores are Attitude Toward School, Mechanics of Study, and Planning and System. In addition to these scores, a verification key is available. The test manual suggests a critical score of 17 and indicates that verification scores below this point usually mean that the student is either attempting to fake an unduly favorable score, is careless in his marking of the test, or has omitted too many items. In this research, pupils below the suggested critical score were eliminated prior to carrying out the analysis.

Unlike the CSMS, the SSHA yields a total score only. The measure contains 75 items, about half dealing with study methods and procedures and half dealing with study attitudes. The test-retest reliability of the total score is reported as .83 for boys and .80 for girls (19), which is comparable to the total score reliablity of .86 for the CSMS (23).

RESULTS

Sample IV

Because of the lack of suitable study habits measures that can be administered at the elementary level, Sample IV pupils were not administered a study habits test until the final year of the study when they were in grade seven. Near the end of grade seven, these pupils were administered the California Study Methods Survey (CSMS). During three years of the four devoted to this research, these pupils were at the elementary level. Thus, the scores obtained in grade seven provide the best evidence available in this project on the relative effectiveness of random grouping with enrichment versus ability grouping with acceleration in developing study methods at the elementary level.

As no suitable study methods measure was available for administration to Sample IV at the beginning of the study, one cannot be certain that initial differences had not already developed between pupils in the two districts prior to the beginning of the research project. The elementary schools from which these samples were drawn were closely comparable, however, and both districts had employed random grouping in self contained classrooms prior to the research. Thus, it seems reasonably safe to assume that the groups were initially similar in terms of study habits. In any event, it would be surprising to find significant study habit differences developing in two such similar school systems by the beginning of grade four.

Results obtained from the seventh grade administration of the California Study Methods Survey to 727 Sample IV pupils may be found in Table 9.

In looking over these results one is impressed by the consistently higher scores of District R pupils. Although it is possible that some of these differences were already present at the beginning of the fourth grade it appears highly probable that at least part of the differences that emerged are attributable to the grouping treatments. District R superior pupils made significantly higher mean scores on all three of the CSMS subtest scores plus the total score as compared with District A superior pupils. All differences between average groups in the two districts were also statistically significant favoring District R. Among slow pupils, the Mechanics of Study mean significantly favored the District R sample at the .01 level, as did the total score. Differences between the slow groups on the other two subtests were not statistically significant.

It is interesting to speculate on possible reasons for the consistent superiority of the study habit scores earned by District R pupils in Sample IV. For the superior pupil, it may be that the enrichment program in effect in District R elementary schools provided more time for independent study than was available in the ability grouped system with its accelerated curriculum. In the process of independent study the superior pupil in District R may have developed the better study methods that are reflected in the CSMS scores.

This explanation, however, would hardly account for the differences in average pupils, because the amount of time the average pupil has to devote to enrichment materials in the heterogeneous classroom is rather limited and the materials available are usually aimed at the extreme groups rather than the average. These pupils, however, may have been influenced by another factor, namely the challenge presented to them by the presence of superior pupils may have motivated the average pupils to set higher goals, and better study habits and attitudes may have resulted. This same factor could also have operated to motivate low ability pupils although one would expect it to be less effective for this group because many of the goals and performance standards set by superior pupils in the heterogeneous class would be unattainable for students of low ability.

One must conclude on the basis of the Sample IV data that ability grouping of the type used in District A does not lead to better study habits for elementary pupils at any ability level. Instead, the data indicate that heterogeneous grouping with enrichment tends to develop better study habits at the elementary level, especially among superior and average pupils.

Sample VI

The California Study Methods Survey was administered to Sample VI pupils near the end of the ninth grade. As was the case with Sample IV, no initial study habits measure was obtained. For the same reasons given for Sample IV, however, there seems to be a sound basis for expecting that little or no initial difference existed between pupils of comparable ability in the two districts.

After eliminating incomplete cases and those who scored less than 17 on the CSMS verification score, a total of 673 complete cases were available for analysis of variance. A summary of the results obtained by comparing pupils of similar ability in District A and District R may be found in Table 10.

In looking over the differences on the CSMS for Sample VI pupils in the two districts the lack of significant differences for superior and slow pupils provide a sharp contrast with the Sample IV results. For both of the samples, District R average pupils were significantly higher on all three subtests and on the total scores. District R superior and slow pupils received significantly higher scores on all three subtests and on the total scores in Sample IV, while none of these differences was significant for Sample VI.

One might wonder why similar differences were not found for Sample VI. Perhaps the most likely explanation is that Sample IV results for the most part reflect changes due to differences in the grouping treatments at the elementary school level; while Sample VI data reflect, for the most part, the effects of the two grouping systems at the junior high school level. At the elementary level the grouping treatments influence the entire school day, while at the junior high-school level, pupils were in ability grouped classrooms for mathematics and science only.

The lack of significant between-treatment differences for superior and slow pupils would suggest that at the junior high school level, enrichment is less effective in developing study habits than at the elementary school level. This might be expected because the self-contained classroom would seem to offer better opportunities for enrichment than the departmentalized class organization employed at the junior high-school level.

Perhaps the best explanation for the consistently higher scores of District R average groups in both Sample IV and Sample VI is that average pupils in heterogeneous classes are challenged to a greater extent than average pupils in classes from which the superior pupils have been removed. It is doubtful if enrichment contributed materially to the better study habits of District R average pupils at either grade level as the enrichment program tends to em-

TABLE 9

MEAN SCORES OF SAMPLE IV SUBGROUPS ON THE CALIFORNIA STUDY METHODS SURVEY

		Means			
Group	N	Attitudes Toward School	Mechanics of Study	Planning and System	Total Score
Superior					
Ability Grouped (Dist. A)	97	35.59	40.25	20.51	96.37
Random Grouped (Dist. R)	202	38.52[a]	41.86[b]	22.23[a]	102.62[a]
Average					
Ability Grouped (Dist. A)	148	31.57	36.32	19.22	87.10
Random Grouped (Dist. R)	157	34.43[a]	38.30[a]	20.57[b]	93.29[a]
Slow					
Ability Grouped (Dist. A)	42	30.33	32.28	19.14	81.75
Random Grouped (Dist. R)	81	31.39	36.42[a]	19.74	87.58[a]
Total	727				

[a] This indicates a difference significant at the .01 level favoring the indicated group as compared with the comparable ability level in the other district.
[b] This indicated a difference significant at the .05 level.

TABLE 10

MEAN SCORES OF SAMPLE VI SUBGROUPS ON THE CALIFORNIA STUDY METHODS SURVEY

		Means			
Group	N	Attitudes Toward School	Mechanics of Study	Planning and System	Total Score
Superior					
Ability Grouped (Dist. A)	69	35.35	43.01	20.33	98.91
Random Grouped (Dist. R)	107	36.03	43.17	19.66	98.91
Average					
Ability Grouped (Dist. A)	226	30.21	38.12	18.13	86.37
Random Grouped (Dist. R)	165	33.91[a]	40.89[a]	19.10[b]	93.90[a]
Slow					
Ability Grouped (Dist. A)	29	29.79	34.48	18.07	82.34
Random Grouped (Dist. R)	77	29.43	36.70	17.93	84.06
Total	673				

[a] This indicates a difference significant at the .01 level favoring the indicated group as compared with the comparable ability level in the other district.
[b] This indicated a difference significant at the .05 level.

phasize adjustment of the curriculum for the superior and slow pupils and as a rule, little adjustment is provided for pupils of average ability.

Sample VIII

The Brown Holtzman Survey of Study Habits and Attitudes (SSHA) was administered to Sample VIII and Sample IX pupils. As both of these samples were tested near the end of junior high school and again near the end of high school, they provide evidence primarily upon the relative effectiveness of ability grouping versus random grouping at the high-school level. The administration of the SSHA on at least two occasions permitted adjustment for initial differences, using covariance analysis.

As the SSHA is not designed for administration prior to grade nine, Sample VIII pupils were not tested on this measure during the first year of the study. They were, however, tested during their ninth grade year and again during their eleventh grade year. Complete data required for the analysis of covariance were available for 428 pupils in Sample VIII.

The adjusted eleventh grade means for superior pupils in the two districts were very similar, being less than one raw score point apart. Slow pupils were also found to be very similar. Average pupils appear to have fared somewhat better in District R as indicated by a significant difference (.01 level) in adjusted final means. In terms of the published norms, District A average pupils were at the 35th percentile while District R pupils were at the 50th percentile (19).

We may conclude from the Sample VIII data that the high-school pupil of average ability develops better study habits and attitudes in heterogeneous classes, probably because of the greater challenge or motivation provided him by the presence of superior pupils in his classes. On the other hand, there seems nothing to choose between the two grouping systems in terms of their effectiveness in developing better study habits in superior or slow pupils.

Sample IX

Sample IX pupils were studied from the beginning of the ninth grade through the end of the twelfth grade. This sample was administered the SSHA on three occasions during the study: near the end of the ninth grade, near the end of the tenth grade and near the end of the twelfth grade. Complete data were available for 214 pupils. The twelfth grade means were adjusted for initial differences using analysis of covariance.

The adjusted twelfth grade means for pupils of comparable ability in the two districts were very similar with slightly more than one raw score point separating the slow groups and an even smaller amount separating the average groups and the superior groups. None of the between-treatment differences approached statistical significance.

In terms of gains in study habits over the period from the ninth grade to the twelfth grade, superior pupils in both districts made reasonably large gains. The average groups and slow groups, however, gained very little, earning virtually the same scores on the test in the twelfth grade that they had earned in the ninth grade.

The Sample IX results seem to suggest that neither the ability grouping system nor the random grouping system as practiced by the districts involved in the Utah study makes any unique contribution to the study habits of high school students.

SUMMARY AND CONCLUSIONS

The Sample IV results on the CSMS showed all of the total score differences and seven of the nine subtest differences between comparable groups in Districts A and R to be statistically significant. Differences were generally largest between pupils of superior ability. These results lead to the conclusion that pupils in a random grouping situation consistently develop better study methods during the elementary school years than pupils in an ability grouping situation.

Sample VI average pupils in District R received consistently higher scores on the CSMS than comparable District A pupils. No significant between-treatment differences were present for slow or superior pupils. Sample VIII average pupils in District R received higher SSHA scores. No significant between-treatment differences emerged from the Sample IX data. It may be concluded that if any advantage accrues from heterogeneous grouping at the secondary level, it is for average pupils, who made significantly greater gains in study habits in Samples VI and VIII. Although this difference was not found in Sample IX, the larger number of subjects in the other samples would still justify this conclusion.

CHAPTER 6

NON-COGNITIVE VARIABLES

THE VAST majority of studies published in the area of ability grouping have been concerned with achievement. Yet, the critics of ability grouping almost never base their criticisms on achievement, but instead bring up the possible effects of ability grouping upon the non-cognitive aspects of the pupils' school experience. They question whether children can experience normal social development in ability grouped classes. They suggest that bright pupils will become snobs, and slow pupils will develop inferiority feelings. They ask if the pupil's self-concept will be damaged, and whether negative attitudes will develop. In this section, we will review the rather limited research pertinent to the non-cognitive effects of grouping and will then examine the findings of the Utah study that relate to some of the aforementioned questions.

The non-cognitive variables covered in this section of the Utah study and the samples tested with each are summarized in Table 11.

PREVIOUS RESEARCH RELATING ABILITY GROUPING TO NON-COGNITIVE VARIABLES

Social Aspects of Ability Grouping

Ability grouping has probably been criticized more frequently in regard to its social aspects than it has in regard to academic achievement. Charges that ability grouping causes social maladjustment are typical of the tirades loosed in opinion articles by Tonsor (113), Bettelheim (8), Cavanaugh (25), and Ginsberg (41). Lawson (70), conversely, contended that homogeneous grouping would not cause maladjustment in the child. He felt that there is a natural tendency for children to gravitate toward contacts with other children of similar intellectual levels. When a child is with others with whom he cannot compete he meets constant failure that is apt to cause maladjustment. Goldberg (42) also felt that placing youngsters into narrow-range ability groups would not bring about poor social or personal adjustment. McGaughy (82), in listing opinions for and against ability grouping, pointed out that students with similar rates of speed are happier together and enjoy school life more, but also that children in slow groups are marked as "dullards" or "dumbbells".

It is readily apparent that there is no uniformity of opinion regarding the effect of ability grouping on the social development of the pupil. There is evidence available indicating that the claim that students are more likely to develop undesirable social attitudes in ability grouped sections than in random grouped sections is not well founded. Studies by Goldberg, Passow and Lorge (46), Goldworth (47), Rochfort (97), and Bell (7) showed little difference in the social or personal attitudes of students growing out of the two plans of grouping. Barbe (4), on the basis of a questionnaire mailed out to 456 students of special high school classes for gifted children, found that the majority of the respondents reported that having been members of these special classes aided them in making a good adjustment. This contradicted the usual complaint about homogeneous grouping to the effect that such children have a more difficult time adjusting to other groups. Such data, however, must be interpreted with caution because of the possibility that the responding group represented a biased sample.

Jackson (63), working with third and fifth grade pupils, reported findings generally favorable to ability grouping. Not only were the schools bothered with only minor behavior problems (no mention was made of the extent of behavior problems earlier), but these schools found that in ability grouped sections truancy almost disappeared.

Martinson's (78) study presents some interesting evidence on sociometric choices of superior pupils. Her experimental groups consisted of 929 boys and girls from every type of public school designed to represent the entire school population of California. The mean IQ of the sample, based on the revised Stanford-Binet scale, was 142. 6. The subjects were matched by pairs with a control group on the basis of chronological age, IQ, sex, and socioeconomic status. The programs in which the subjects participated were planned within the general areas of enrichment within the classroom, acceleration, and special groupings. In these areas a total of 17 programs at various grade levels were evaluated.

It was found that first graders in the experimental programs showed highly significant gains in friendship choices, while the control gifted subjects showed no change, either positive or negative. Again at the fifth and sixth grade levels, the total experimental group of 191 subjects showed highly significant gains in social status on the basis of responses by their peers in regular classroom situations, while their matched controls showed no change. Apparently the gifted child's classmates did not regard him as a less desirable friend because special provisions were made for him. Gold-

TABLE 11

SAMPLES ADMINISTERED NON-COGNITIVE MEASURES IN THE UTAH STUDY

Measure	Sample and Grade Level to Which Each Measure was Administered				
	Sample IV	Sample VI	Sample VII	Sample VIII	Sample IX
1. Three criterion Sociometric Choice Measure	4th, 5th, 6th	6th			
2. USU School Inventory (attitudes toward peers, teachers, and school)	6th				
3. SRA Junior and Youth Inventories (Problem check lists)	5th	7th	8th	9th	
4. Index of Adjustment and Values (four scores related to self-concept)	5th, 6th, 7th	7th, 8th, 9th	8th, 9th	9th, 10th	9th, 11th
5. California Test of Personality (three subtests used)	5th				
6. Inventory of Factors GAMIN (two factors used, ascendency and inferiority)		7th	8th	9th	
7. Objective-Analytic Personality Test Factors					
UI – 16 -- Competent Assertiveness	5th				
UI – 24 -- Anxiety to Achieve		7th			
8. California Psychological Inventory (all 18 scores obtained)		8th	9th		
9. Individually administered TAT type measure (three scores – aggression, depression, and inferiority feelings)	5th (slow)				
10. Groups administered TAT type measure	6th				

berg (43: 51) supported these findings in her statement that "There is no evidence to support the notion that grouping has adverse effects on the social or personal attitudes or behavior of children."

Drews' study (31), was concerned with the effects of homogeneous and heterogeneous grouping on the performance of superior, average and slow pupils in ninth grade English classes. Her data provide some valuable evidence on the effects of grouping upon sociometric status. In her study, six three-choice sociometric questions were used. Two of these questions required pupil choices on social variables while the other four were concerned with intellectual variables. Pupil choices were analyzed by question and also composite peer choice scores in the social and intellectual categories were calculated. The pupil's choices were limited to individuals in his English class, which places considerable restriction on choices in the departmentalized junior high school situation. In the heterogeneous classes, Drews found that for each of the six sociometric criteria, superior students were chosen more frequently than average students and slow students were least often selected. Although marked differences occurred in both the social and intellectual categories, the differences, as might be expected, were greater in the intellectual area. Drews also compared the distributions of sociometric choice scores obtained by superior, average and slow pupils in her homogeneous and heterogeneous classes. Comparisons of mean scores in the two types of classes would, of course, be spurious, because in the homogeneous class the pupil has no choice but to nominate persons of the same ability level. The distribution of choice scores, however, can be meaningful because if a leadership vacuum exists in the average and slow homogeneous classes as has sometimes been suggested, none of the average and slow pupils would receive enough sociometric choice nominations to be classified as "stars". She found that although a larger percentage of superior pupils received high numbers of nominations in both grouping systems, the differences between ability levels were much greater in the heterogeneous classes. On the intellectual choice questions, if we take a sociometric score of 12 as the cut off point, we find that 54 percent of the superior pupils in heterogeneous groups obtained more than 12 nominations while no slow pupils in the heterogeneous groups reached this level. In contrast, 30 percent of the superior pupils in homogeneous groups received more than 12 nominations while 28 percent of homogeneously grouped slow pupils reached this point. On the social questions, approximately 33 percent of the superior pupils in heterogeneous groups received more than 12 nominations with no slow pupils receiving this number. In homogeneous groups, 30 percent of the superior pupils, and 18 percent of the slow pupils, received 12 or more nominations. These data would indicate that although a smaller proportion of slow than superior pupils in either treatment gained positions of social recognition, the slow pupil had almost no chance to gain such status in the heterogeneous classes, while his chances were nearly as good as the superior pupil in the homogeneous classes. If Drews' study had extended for more than one year, it seems likely that the distributions of sociometric choice scores in the homogeneous classes would have become even more similar as slow pupils gained more experience in positions of social leadership. Many psychologists have stressed the importance of peer acceptance during the adolescence period. In view of this, Drews' evidence would seem to provide strong support for ability grouping at the junior high school level.

Mann's research (74) concluded that there was little real friendship between superior and average or below average students in heterogeneously grouped classrooms. In this study 67 gifted pupils in a total sample of 280 fifth and sixth graders were placed in regular classes for half of their work and in special workshop rooms with other gifted children for the other half. On the basis of sociometric choices, there were more acceptances and rejections of the gifted by the gifted than by the typical children. Typical children, as a group, tended to accept and reject more typical children than workshop children. These findings suggest that friendships were based to a considerable degree upon ability level. It appears a safe assumption that students will select as friends those persons with whom they associate more, especially at the elementary school level. With this in mind one might speculate as to whether the sociometric results of Mann's study were due to similar mental ability or to the greater contact among the gifted children than between the gifted and typical groups. This increased contact would serve to increase the number of sociometric choices regardless of the basis used for separating the children into groups, thus casting serious doubt on the validity of Mann's conclusions.

The literature concerning the social aspects of ability grouping includes at least some evidence to support any stand one might care to take. Most of the research studies concluded that the grouping programs had no adverse effect on the social adjustment of the students. Some evidence was presented that indicated an improved social relationship in ability grouped classes, while other evidence supported the critics of ability grouping by indicating that the development of undesirable attitudes resulted from such grouping programs.

Pupil Attitudes and Grouping

Almost no research has been reported that is directly pertinent to pupil attitudes toward school in ability grouped and random grouped classrooms. Some data in this area were collected by Simpson and Martinson (105). The reader will recall that this study was concerned with the effectiveness of special programs for gifted children in California. Rating scales were developed for use by pupils, teachers, and parents that dealt with a number of areas somewhat related to pupil attitudes. Among the variables covered on these questionnaires were included interest in school and rapport with teachers. These data, however, were not collected for the control groups so that comparisons were not possible for most of the samples involved in the study. The fifth and sixth grade experimental programs, however, included both an enrichment program and a special class program. The first of these programs employed enrichment for superior pupils in the regular classroom, and the second involved a special class for superior pupils. These two programs were quite similar to the random grouping and ability grouping systems in Districts A and R of the Utah study. Rating scale data from pupils and parents were collected on these two groups at the beginning and end of the one year study, and percentage of gains was calculated. The rating scale items on interest in school, motivation towards learning, and intellectual curiosity were classified together as "intellectual stimulus" by the investigators. Ratings by pupils and their parents in the special class situation indicated an average pupil gain on these items of about 80 percent over the year of the study. Similar ratings for the enrichment program indicated gains of about 50 percent. On another pertinent variable, "relationships with the teacher", ratings by special class pupils and their parents showed a gain of about 45 percent as compared with a gain of approximately 35 percent for the enrichment program. Teacher ratings were not available for the special class groups. As each of the rating scales included only a single item for each area covered, these data are subject to serious limitations in reliability. The validity of a direct rating scale when used by presumably untrained raters must also be questioned.

Goldberg and her associates reported that in their study of the effect of ability grouping per se, comparisons between pupils in homogeneous and heterogeneous classes on attitudes towards school and attitudes towards more or less able peers yielded results that were largely negative. [6]

A study of Passow, Goldberg, and Link (89) compared four groups of gifted junior high school pupils in mathematics achievement from the beginning of the seventh to the end of the ninth grade. Two of these groups participated in an accelerated program, one in an enrichment program and there was one control group. In addition to measuring achievement differences, this study administered an attitude inventory at the beginning of the seventh and end of the ninth grade. This inventory contained items in six attitude areas related to mathematics. The groups were closely comparable on the initial attitude test. Analysis of the total score of the attitude inventory administered at the end of the ninth grade showed the two classes in the accelerated program to have similar means (60.2 and 63.6). These means were significantly higher than those of pupils in the enrichment group (53.6) and the control group (53.7). These data suggest that acceleration develops considerably more favorable attitudes toward mathematics than enrichment among groups of homogeneously grouped gifted pupils. A limitation of this study is that only one classroom was involved in each of the four grouping treatments. Four different teachers taught these classes. Thus, it is very possible that some of the differences obtained were due to teacher differences rather than to treatment differences.

Drews' study (31) of homogeneous and heterogeneous grouping in ninth grade English classes was also concerned with pupil attitudes. Interest in school was obtained by asking pupils to check "yes" or "no" in answer to the question, "Is school work interesting to you?" These data were obtained at the beginning and end of the one year study, and indicate that the two grouping treatments had no effect on the pupils' stated interest in school. The use of a single direct question to measure pupil interest raises doubt about the reliability and validity of these data.

Attitude towards English as a school subject was also obtained at the beginning and end of the research project. The results showed that attitudes toward English improved for all groups excepting the heterogeneous slow group between the pre-testing and post-testing. Four questions were asked of pupils in the two treatments to see if differences had developed between the attitudes of pupils in the homogeneous versus the heterogeneous English classes. The results reflected about the same attitudes for average pupils in the two treatments, but for superior and slow pupils the homogeneous classes elicited much more favorable attitudes.

Self-Concept and Ability Grouping

Only four studies have been reported in the literature that are concerned with the relationship between ability grouping and self-concept. The first such study was carried out by Mann (75) with a sample of 102 fifth grade pupils who had been grouped into four ability levels since entrance into the first grade. Grouping was based upon the results of group intelligence tests and reading readiness tests. Sections were labled only by the teachers' names but teachers informally referred to them as "highest", "lowest", "second highest", and "second lowest" groups. Self-concept data used in this study were based on the subjects' anonymous responses to two questions. One of these was, "Tell me which fifth grade you are in", and the other was, "Tell me how you happen to be in this particular fifth grade group rather than some other group." About two-thirds of the lowest and highest groups identified their groups by ability level. A great majority of the two center groups identified their groups by the teacher's name. In response to being asked why they were in a particular group, 25 of the 30 pupils in the highest group answered in positive terms relative to ability or achievement such as "I'm smart", "I make good grades", "We work a little faster", and so on. In the lowest group, 14 of the 18 pupils answered in negative terms relative to ability or achievement such as "I am too dumb", "I am not so smart", "We don't know too much." Mann's study clearly indicates that pupils in an ability grouped situation are aware of the nature of their situation, but presents no direct evidence on the effect of grouping on the student. No control group was employed. It is possible that use of a control group would have revealed that low and high ability pupils in random groups have similar low and high perceptions of their ability and achievement. Furthermore, a questionnaire based on only two items would be expected to have rather low reliability. Mann's study would have been improved considerably by the use of a more extensive measure of self-concept. Other studies such as the work of Luchins and Luchins (72), have also suggested that pupils in an ability grouped situation are aware of their grouping status regardless of efforts made by the school to conceal the grouping level. This awareness, coupled with the associations that the pupil makes with his ability level as indicated by Mann, certainly suggests that ability grouping may have undesirable effects upon the pupil's self-concept.

Drews' research (31) with ability grouped and random grouped classes in ninth grade English also included measures of self-concept. Drews hypothesized that slow pupils would see themselves as relatively more competent in a homogeneous situation while the reverse would be true of superior pupils. Two self-concept measures were employed to test this hypothesis. The Ability Self-Concept Rating consisted of a single question asking the student to compare his ability with that of his classmates and rate himself as above average, about average, or below average. The other measure used was the 20 item Concept of Self-as-a-Learner Scale which Drews developed from Bills' Index of Adjustment and Values.

The Ability Self-Concept Rating was administered at the beginning and end of the experimental year. Mean scores were analyzed using the t-test to determine if significant differences existed between homogeneous and heterogeneous classes. Only one significant difference was found on the pretest analysis – favoring slow pupils in the homogeneous sample. At the end of the one year experiment, between-treatment comparisons revealed that superior pupils in heterogeneous groups and slow pupils in homogeneous groups made significantly higher scores on this measure.

Drews administered the Concept-of-Self-as-a Learner Scale at the end of the study only. She found that although the heterogeneously grouped superior pupils obtained higher mean scores, the differences were not significant. Only one significant between-treatment difference occurred, favoring the slow boys in the homogeneous groups.

An important study by Goldberg and her associates has not been completed at the time of this writing. She reports, however, that self-attitudes were significantly affected by grouping.

> In narrow range groups the gifted children showed a significant decrease in their self-appraisals which were considerably inflated to begin with. On the other hand, the slow children in narrow-range groups showed increased self-appraisals which brought them more nearly to the level of the average. The gifted showed considerable increases in self-appraisal while the slow decreased significantly. Depending upon your point of view, effects of grouping on self-attitudes may be perceived as desirable or undesirable. The effects of grouping were even more startling with regard to self-aspirations. There, the broad range group, and especially the presence of gifted children tended to act as a damper on the aspirational ceilings of the slower pupils.[7]

A portion of the California study of gifted children (105) showed that eighth grade pupils in special classes made greater gains on Self-Acceptance as measured by the California Psychological Inventory than did pupils of comparable IQ, who were assigned to heterogeneous control classes. These findings agree with those reported by Goldberg but disagree with Drews (31). Perhaps these findings differ because Drews' study made special curriculum adjustments for the heterogeneous groups while the other two studies did not.

Personality and Ability Grouping

The controversy over ability groupings' effect upon personality is to a large extent based on opinion and philosophical orientation. A variety of behavioral scientists including Bettelheim (8), Hamalainen (56), Tonsor (113), and Gowan (50) have stated their opinions of the effects of ability grouping on the personality of the pupil. Most educators also have strong feelings on this topic, but when we look for research evidence there is very little. Let us now review the limited evidence available on the effects of ability grouping on personality.

In an individual interview study, Luchins and Luchins (72) questioned 190 fourth , fifth and sixth grade children about their attitudes toward the ability grouping practiced in their school. Every pupil was aware of the significance of the superscript, "1" (bright), "2" (dull), "3" (average) attached to his class. Practically all students in the "1" class preferred their class and would choose to remain in it even if they did not like the teacher. Among the reasons for choosing the "1" class were: they belonged there, they would be bored, they did not want to mingle with or be educated with "dumbbells". When asked which class they would rather be in, children in the average and dull classes chose "1" class because there would be stigma attached to being in "2" class which "has a reputation you can't get rid of", their parents would be proud of them, the teacher would not be so strict, they would have more fun, they would learn more, and besides it paid to take the chance since they could always go back to "2" class. The few students in the bright and average classes who selected the "2" class did so because they might get better grades or they could be more outstanding. Those in the "2" class who selected to remain said it was best for them, they would be out of place in "1" class, and they wouldn't be allowed in "1" class anyway, so why should they pick it? Luchins and Luchins concluded that the prestige value of "1" class membership was its chief attraction, that it caused the able members to become snobbish, that the grouping situation in the school caused stigma to be attached to the "2" class membership, and that homogeneous grouping created a caste system in the school. Since the subjects in Luchins and Luchins' study were not compared with any controls, it is impossible to conclude definitely that the findings were attributable to the grouping situation, although this appears to be the case. Students' feelings, like those reported in Luchins and Luchins' study, appear to be evident to school administrators. A survey by Saville (103) revealed that a number of school officials considered the "stigma" attached to pupils involved in ability grouping to be a major problem.

Mathias (81), using the Rorschach and TAT to measure aggression and depression, found that the slow and bright children had significantly more aggression responses (beyond the .01 level of significance) than average pupils in heterogeneous classes. He concluded this was due to frustration they experienced in a school program aimed primarily at the average pupil.

Simpson and Martinson (105) administered the California Psychological Inventory to their special class groups and control groups at the eighth-grade and high-school levels. Subjects involved in the eighth-grade phase of the Simpson and Martinson study included 58 boys and 57 girls in the special classes and 36 boys and 20 girls in the control group. It will be recalled that this study was concerned with superior pupils only. The special class program was similar to ability grouping except that it involved superior pupils only. The CPI was administered to the special class and control groups at the beginning and end of the one-year study and gains in the subtest scores were tested for significance.

The special class groups made significant gains in 19 instances and significant losses in three instances on the CPI. The control group pupils made significant gains in nine instances and significant losses in eight instances on the CPI.

This study did not present direct statistical comparisons to determine whether the experimental groups made significantly greater gains than the control groups. Examination of the eighth-grade data available for boys would indicate, however, that the special class group probably made significantly greater gains on Self-Acceptance than the control group. Similar comparisons for the eighth-grade girls appear to indicate that the special class girls made a significantly greater relative gain in Self-Acceptance and Flexibility than the control group.

The data available for the high-school boys in the special class and control groups would indicate that the special class sample made a significantly greater gain in Social Presence and Tolerance.

High-school girls in the control group appear to have made significantly greater gains than the special class group in Social Presence.

CHAPTER 7

SOCIOMETRIC CHOICE[8]

PEOPLE ARE an important part of the environment to which individuals react in the processes of developing and learning. Wrightstone, Justman and Robbins (124: 199) have asserted that "The development of one's personality is a product of the interaction of the individual with other people." According to Otto (88) it is in these person-to-person relationships in various sizes and types of groups that many social needs are met and personality development has its roots. The need to belong to a group is a powerful force. Some psychologists claim that this need is an innate human need while others maintain that it is culturally developed. All agree that, whatever its origin, the need to belong furnishes strong motivation for behavior, for in our society almost every individual attempts to become a member of some group. Cunningham, et al. (28: 108) state that "It is important for the development of most boys and girls of school age that they find belonging in groups of their peers." Dineen and Gary (30: 358) express their agreement with this concept in maintaining that;

> Chief among the factors necessary for developing a full life in a heterogeneous group such as a class of children is a satisfactory friendship pattern. When a child is with others who accept him and respond to him, that is, others with whom he wants to associate, he can contribute more and function better in the group.

Social adjustment has become of major importance in education. The alert teacher is aware of the need of children to establish and maintain emotional and social stability. The relationship which exists between social adjustment and academic achievement has been shown in several recent studies (22, 112, 24, 102). In ability grouping, the matter of social relationships and adjustments has been of primary concern to critics. One of the criticisms most often directed at ability grouping in the public schools is that the children will suffer social maladjustment as a result of being placed in a classroom limited to fellow students of similar intellectual caliber. Such a system results in an artificial social situation that stigmatizes slow students, according to Hamalainen (56). On the other hand, many of those who advocate ability grouping claim that the slow child has an opportunity to be recognized socially in such groups. He is denied such recognition in randomly grouped classes where he must compete with more able students.

The primary purpose of this phase of the Utah study was to investigate the effects of homogeneous versus heterogeneous grouping practices on the sociometric structure of elementary school classrooms. This represented an attempt to determine if systematic differences exist which may be attributable to the type of grouping being practiced.

Sociometric status is defined as the total number of times a pupil was selected on a sociometric choice measure by his classmates as a preferred companion on any one of the three specified activities or situations used as criteria. Mobility of sociometric status may be defined as movement from one sociometric classification to another in consecutive years.[9] Mobility of status over a three year period was studied in this project in an attempt to determine whether the social relationships measured were transitory or relatively stable. Extending the study over a three year period also promised to give some indication of the degree to which increasing exposure to the two grouping treatments and increasing age and maturity, with its accompanying changes in personality characteristics and interests, might affect one's social relationships. Are status and acceptance based on stable personality traits, or as Staker (107) has suggested, are there different traits that are important at different age levels?

A "star" is defined as an individual who receives a large number of choices on a sociometric test, more than could be expected by chance alone. One who receives fewer choices than could be expected by chance, is known as a "neglectee", while those falling between stars and the neglectees are "regulars". An "isolate" is one who receives no choices, while a "rejectee" is one who receives negative choices, if negative choices have been obtained.

Gronlund and Whitney (53) have suggested that social acceptance pervades all aspects of one's life sphere. This would mean that those factors that contribute to one's acceptance by others operate at home, on the job, at school, in informal contacts, or in any situation in which a person may become involved. This idea was supported by Symonds and Jackson (110).

The design of the Utah study provided for the investigation of the effects of three different conditions on measures of sociometric status. These conditions were 1) the grouping treatment (ability grouped or random grouped) 2) ability level, (based on academic achievement) and 3) sex. Comparisons were made between the two districts, among the various ability levels, and between the sexes on the aforementioned dependent variables.

WORKING HYPOTHESES

The following working hypotheses were established regarding the effects of grouping treatment, ability level, and sex on the sociometric variables.

1. It was hypothesized that between-treatment comparisons would reveal fewer stars among District A superior pupils but greater numbers of stars among both average and slow pupils in District A.

2. Furthermore, it was hypothesized that the combined distribution of pupils at all ability levels into the three sociometric categories of "stars", "regulars" and "neglectee-isolates", would not be significantly different for the two districts.

A number of previous studies have shown relationships between sociometric choice scores and achievement or ability in heterogeneous classrooms. When one considers the ability grouping situation, however, where achievement variability is somewhat reduced, it appears probable that the effect upon superior pupils would be to reduce their chances of gaining social recognition. Two changes in the classroom social situation seem possible for average and slow pupils when placed in an ability grouped situation. One of the possible effects of ability grouping could be the development of a leadership vacuum in the average and slow groups resulting in few if any recognized social leaders or stars in these groups. The second alternative would be that, with the superior pupils removed, pupils in average and slow groups who had not previously been recognized in the peer structure would emerge as social leaders or stars. If the first of the two aforementioned alternatives occurred, the overall proportion of stars would be considerably less in District A than District R. On the other hand, if the latter situation occurred, the number of stars in District A would be equal or perhaps greater than in District R, as fewer superior pupils would probably attain the status of stars in the ability grouped class but greater numbers of both average and slow pupils would reach this level of acceptance. Working hypotheses concerned with sociometric choice patterns in the two districts involved in the Utah study were based upon the latter expectation.

3. Regarding level differences within each treatment, it was hypothesized that in the random grouped situation, where previous studies have demonstrated a relationship between status and ability, superior pupils would obtain the more favorable sociometric status patterns (more stars and fewer neglectee-isolates) followed by average and slow pupils in that order. In District A, where ability differences within the classroom were somewhat less, it was hypothesized that sociometric choice patterns for pupils at the three ability levels would not be significantly different. That is, superior, average, and slow pupils would all have about the same distribution of stars, regulars, and neglectee-isolates.

4. It was hypothesized that no significant sex differences would occur in the sociometric choice patterns or mobility patterns of boys and girls of comparable ability.

The degree to which pupils in the two treatments changed from one to another sociometric classification from year to year appeared likely to provide some valuable insights into the relative effect of the two grouping treatments upon the classroom social structure.

5. It was hypothesized that the overall patterns for average and slow pupils would tend to improve from year to year in the District A classrooms while staying about the same in District R classrooms. This hypothesis was based on the assumption that many average and slow pupils would require a certain amount of learning and experience before they would be able to move out of the background and assume positions of social leadership in ability grouped classes. Thus, although it might be expected that some average and slow pupils would move into positions of social leadership very quickly, it seemed likely that many who had potential for social leadership would require a longer experience before emerging as stars. This lag could result in a temporary leadership vacuum in average and slow sections in District A.

A final goal of this phase of the project was to determine whether significant relationships existed between sociometric status and a number of other variables for which data were collected in other phases of the Utah study.

6. It was hypothesized that sociometric status would correlate positively with ability and achievement for District R pupils but that these variables would not be significantly correlated for District A pupils.

7. It was hypothesized that a positive relationship would exist in both districts between sociometric status and favorable attitudes toward the school environment.

8. Positive relationships were also hypothesized between sociometric status, and self concept variables as measured by the Index of Adjustment and Values.

PROCEDURE

Sample

Sample VI pupils were administered the sociometric measures only once – in the sixth grade – and therefore, this sample was used only in the study of sociometric patterns. Sample IV pupils were measured in the fourth, fifth, and sixth grades, and data from this sample were used to study sociometric status patterns mobility of sociometric status, and relationships between sociometric status and other variables. Over the three years that sociometric data were collected on Sample IV, the number of pupils in District R ranged from 496 to 702, and in District A from 292 to 455. There were 537 Sample VI pupils measured in District R and 425 in District A during the single year that sociometric data were collected on this sample.

Measures Used

A near-sociometric measure of the usual partial-rank-order type was developed for use in this study. Three criteria were employed, each with five positive choices. Each subject was asked to indicate by placing a check mark beside names on a roster of his classmates those whom he preferred under each criterion. The three criteria were: 1) the five children in the class who are your best friends; 2) the five children with whom you prefer to study; and 3) the five children you would most like to have with you if you were transferred to another classroom.

Measures with which the sociometric data were compared included the Bills Index of Adjustment and Values, which was used to measure the self-concepts held by the students; the Sequential Tests of Educational Progress, that were used to obtain a measure of achievement; the California Test of Mental Maturity, and the USU School Inventory, a locally developed attitude scale, measuring attitudes of students toward school, teachers, and peers.

Each pupil's sociometric status score was obtained by summing the choices he received on the three criteria. After obtaining a sociometric choice score for each pupil, the pupils were classified as stars, regulars, neglectees, or isolates based upon choice limits established by Bronfenbrenner (17). Category limits using this system were as follows:

“stars” -------- received 22 or more nominations
“regulars” ----- received 10 to 21 nominations
“neglectees” ---- received 1 to 9 nominations
“isolates” ------ received no nominations

For purposes of the analysis, however, neglectees and isolates were combined. An extremely small number of isolates were obtained in the three-criterion five-choice sociometric measure.

Scores on sociometric status were obtained in this manner for each of the three years. Data that were used to relate sociometric scores to mental ability, academic achievement, attitudes, and concept of self were gathered during the third year of the study when the subjects were in the sixth grade.

Method of Analysis

The study was concerned primarily with the effects of the two grouping treatments, ability level, and sex on sociometric status. The various combinations of these three variables provide 12 groups that could be compared on each dependent variable. These comparisons were tested for significance using either single classification analysis of variance or chi square contingency tables.

RESULTS

Between-Treatment Differences

It will be recalled that two possible effects of ability grouping on sociometric choice patterns were discussed in the working hypotheses. One of these was that ability grouping would create a leadership vacuum in ability grouped sections made up of average and slow pupils, thus reducing the overall proportion of stars in District A. In order to check this possibility, the overall distribution of stars, regulars, and neglectee-isolates for District A and R were compared. Four overall comparisons were made, one for each year that Sample IV pupils were tested; and one for Sample VI pupils. Only one of these comparisons showed a significant difference between the ability grouped and random grouped subjects. This difference, between Sample IV pupils in District A and R during the first year of study revealed a smaller number of stars and a greater number of regular pupils in District A than would occur by chance. These results suggest that during the first year of the ability grouping situation, some leadership vacuum was present in regular and slow groups. A similar trend in the overall sociometric choice pattern also occurred during the first year of the study for Sample VI. The trend, however, did not reach statistical significance for this sample and suggests that sixth grade pupils adapt more rapidly to the ability grouped situation than do fourth-grade pupils (see Table 12). No such trend, however, existed for Sample IV pupils during the second and third years of the study. Overall comparisons for these years indicate a slightly larger number of stars in District A than in District R. These data would suggest that once the child has adapted to his new role in the ability grouped class, the so-called leadership vacuum quickly disappears, and pupils of average and low ability emerge as stars.

In addition to these overall between-treatment comparisons, the sociometric choice patterns of boys and girls at each ability level in District A were compared to those of their counterparts in District R. The results of these comparisons for the three years of the study are shown in Table 12. The reader will recall that our working hypothesis predicted a smaller number of stars among superior pupils in District A. This trend was, in fact, evident in the eight comparisons between superior pupils in the two districts. It will be noted that three of these differences reached statistical significance and one additional approached significance. The three non-significant differences also followed the same trend. In an ability grouping situation, supe-

TABLE 12

DIFFERENCES BETWEEN DIFFERENT SUB-GROUPS IN THE FOUR REPLICATIONS BASED UPON CHI-SQUARE CONTINGENCY TABLES

Groups[b]	Purpose of Comparison	1958-59 4th Grade Sample IV	1959-60 5th Grade Sample IV	1960-61 6th Grade Sample IV	1958-59 6th Grade Sample VI
Between-Treatment Differences	Overall Differences: To check overall proportion of sociometric classification between groups.				
1. All subjects in District R versus all subjects in District A.		6.59[a]	N. S.	N. S.	N. S.
2. Superior Boys, R versus A	These 6 comparisons are between comparable sub-groups in terms of sex and ability in District A versus District R.	N. S.	N. S.	N. S.	6.52
3. Superior Girls, R versus A		8.56	A. S.	N. S.	6.04
4. Average Boys, A versus R		A. S.[c]	8.53	N. S.	A. S.
5. Average Girls, A versus R		N. S.	N. S.	A. S.	N. S.
6. Slow Boys, A versus R		N. S.	12.87	A. S.	N. S.
7. Slow Girls, A versus R		N. S.	9.25	N. S.	7.38

[a] Chi-square of 5.99 significant at .05 level; 9.21 significant at .01 level; 13.8 significant at .001 level

[b] Group having more desirable distribution of sociometric classifications listed first when significant differences are present.

[c] A. S. indicates that difference approached significance (significance level between .05 and .10).

rior pupils, being placed together, must compete among themselves for a smaller number of positions of social leadership than would be available to them in random grouped classrooms. It is noteworthy that, although ability grouping reduced the superior pupil's chances of reaching the star classification, it did not increase the liklihood of his being classified as a neglectee-isolate.

It was hypothesized that larger numbers of stars would be present among average pupils in ability grouped classrooms because of the positions of social leadership opened to these pupils by removal of the superior pupils from the classroom. It will be noted in Table 12 that one of the eight between-treatment differences for average pupils was statistically significant and two others approached significance. These differences plus the remaining four non-significant differences were all in the hypothesized direction.

It was also hypothesized that a greater number of stars would emerge among slow pupils in District A because of the removal of superior and average pupils who might be expected to dominate the social situation in a heterogeneous classroom. It will be noted in Table 12 that three of the eight comparisons between slow pupils in the two districts were statistically significant while a fourth approached significance. Examination of the eight chi-square contingency tables for slow pupils revealed that significant differences in all cases were due to a greater number of stars and a smaller number of neglectee-isolates in District A as compared with District R. Contingency tables not yielding statistically significant chi-squares were also examined and it was found that the above trend also occurred in these tables. It is noteworthy that between-treatment differences between Sample IV slow pupils during the first year were not significant for either boys or girls, again suggesting that slow pupils require a certain amount of time for adjustment before they emerge as leaders in the ability grouped situation.

The total pattern revealed in the between treatment comparisons is strikingly consistent. This pattern seems to indicate that average and slow pupils generally have a far better opportunity to attain positions of social leadership and are far less likely to be classified as neglectee-isolates in an ability grouped situation than in a random grouped situation. Superior pupils generally have less chance of emerging as stars in the ability grouped classroom. Examination of the contingency tables indicate, however, that superior pupils do not more frequently become neglectee-isolates in the ability grouped classroom.

Part of the aforementioned differences are probably an artifact of the ability grouping situation, that is, that pupils in ability grouped classrooms have no one to choose except their classmates at the same ability level. It should be noted, however, that this situation would not cause slow or average pupils to emerge as stars, unless they had in fact gained the level of social recognition necessary for star status. The ability grouped situation could have produced a vacuum in social leadership in which no one would have emerged as a star. The fact that such a vacuum did not exist (except in the fourth grade) indicates that slow and average pupils made real gains in social leadership in the ability grouped classes.

Level Differences

It was hypothesized that no significant differences between District A pupils at different ability levels would be found. This hypothesis was based on the assumption that in the ability grouped situation pupils have about an equal chance of attaining favorable sociometric positions regardless of the ability level of the group in which they have been placed. It was thought that some differences between superior and slow pupils might emerge in District A during the first year of the study because of the possibility that slow pupils would be unable to exploit immediately the better opportunities for social status available to them in ability grouped classes. Of the 24 comparisons between pupils of different ability levels in District A, only one reached statistical significance at the .05 level. The significant difference was between the sociometric patterns of superior and slow boys and reflected more favorable patterns for superior boys. This difference again suggests that slow pupils are not capable of filling immediately the leadership vacuum that is created when ability grouped classes are established. These data also seem to indicate in a remarkably consistent manner that in the ability grouped situation pupils have an equal opportunity to attain a favored sociometric position regardless of their ability level.

It was hypothesized that significant differences would occur between pupils of different ability levels in District R. This hypothesis was based largely on previous research indicating that in the heterogeneous class the superior pupil has a greater chance of gaining peer group recognition than his less capable contemporaries. It was also hypothesized that in comparisons between average and slow pupils, average pupils would have the more favorable sociometric choice pattern but that these differences would generally be smaller than those involving superior pupils. As superior pupils in the heterogeneous classrooms tend to monopolize positions of social leadership, the lack of opportunity for both average and slow pupils to emerge as stars would reduce the differences between these two groups. A total of 24 chi-square comparisons were made between District R pupils of different ability levels. The eight comparisons between superior and slow pupils were all statistically significant beyond the .01 level with seven of the eight significant beyond the .001 level. Of the eight comparisons between superior and average pupils in District R, three were significant beyond the .001 level, three were significant beyond the .01 level, one approached significance, and one was non-significant. The two non-significant differences were consistent in direction with the significant differences favoring the superior pupils. These data seem to present overwhelming evidence that the superior pupil is in a highly favored sociometric position in the random grouped classroom as compared with average or slow pupils.

Of the eight comparisons between average and slow pupils, only two were statistically significant. Both the significant and non-significant differences generally favored average pupils. These data reflect a much larger gap between superior pupils on one hand and average and slow pupils on the other hand, than is found between the average and slow groups in terms of their success in obtaining favorable sociometric classifications in the random grouped classroom.

Sex Differences

Comparisons were also made between the distributions of stars, regulars, and neglectee-isolates for boys and girls of comparable ability in each district. It was hypothesized that no significant sex difference would emerge. For District A, none of the 12 between-sex comparisons yielded significant chi-square values. For District R, two of the between-sex comparisons reached statistical significance. Both of these were between superior boys and girls, and in both cases, the girls obtained the more favorable distribution of sociometric choices.

Mobility in Sociometric Status

In order to determine mobility of sociometric choice status over the three year period of study, the sociometric classification (star, regular, or neglectee-isolate) was determined for each pupil during each of the three years of the study. Pupils were then classified into three groups: those moving to a lower sociometric category, and those staying to a higher sociometric category, those moving to a lower sociometric category, and those stay-

TABLE 13

BETWEEN-TREATMENT DIFFERENCES IN SOCIOMETRIC STATUS MOBILITY AT EACH ABILITY LEVEL

	Significant Chi Square Values		
Group	Year 1-2 1958-59 1959-60 t-value	Year 2-3 1959-60 1960-61 t-value	Year 1-3 1958-59 1960-61 t-value
Superior pupils	N. S	N. S.	N. S.
Average pupils	A[a]10.97[b]	N. S.	N. S.
Slow pupils	A-12.39[b]	N. S.	A-10.48[b]
Pupils moving downward in ability level	A-11.67[b]	N. S.	N. S.
Pupils moving upward in ability level	R- 6.46[c]	N. S.	R-10.43[b]

[a] Indicates that District A sample made a greater proportion of favorable changes.
[b] Significant at .01 level of confidence.
[c] Significant at .05 level of confidence.

ing the same from one year to the next. Comparisons were made between the first and second years, second and third years, and first and third years. After mobility data had been obtained for each subgroup, comparisons between the patterns of change in sociometric classification between different subgroups were made. Chi-square contingency tables were used to determine the direction and level of significance of between-treatment, level, and sex differences in sociometric choice mobility.

Between-Treatment Differences

Between-treatment comparisons in mobility first were made between the total samples in the two districts. The difference in sociometric status mobility between Districts A and R from the first to the second year was statistically significant beyond the .001 level (chi-square of 14.94). Differences in mobility patterns between years two and three and years one and three were not statistically significant. Examination of the significant contingency table indicated that District A pupils were more mobile during this period. This mobility was in an upward direction for the most part and reflects the marked readjustments in sociometric status that occur when an ability grouping system is adopted. Differences between the second and third year were much smaller. This suggests that by the second year, adjustments in sociometric status that were brought about as a result of introducing the ability grouping system had largely taken place and the new status pattern that had emerged might be expected to remain relatively stable.

In order to get additional insight into between-treatment differences in sociometric mobility, these data were analyzed further by comparing the sociometric status of pupils of the same ability level under the two grouping systems using chi-square contingency tables. Because of a general absence of sex differences on this variable, boys and girls were combined for this analysis. In addition to comparing superior, average, and slow pupils in the two districts, pupils in each district who moved downward or upward in ability level were also compared to determine whether such movement in ability classification was accompanied by changes in sociometric status (see Table 13).

Comparison of pupils of superior ability in the two districts yielded no significant differences in mobility of sociometric status. There was a slight trend evident between years one and two for District A superior pupils to make more status gains and less status losses than were made by District R superior pupils during this same period. This may reflect a tendency for superior pupils to regain some of the sociometric status they initially lose in an ability grouped situation.

Comparisons between the sociometric status mobility patterns of average pupils show District A pupils to be significantly more mobile between years one and two than comparable District R pupils. This mobility was generally upward and reflects an improvement of the average pupils' status during the second year of their experience in the ability grouped situation. The difference between pupils of low ability in the two districts in mobility of sociometric status also reached statistical significance in comparisons between years one and two. These data seem to support the conclusion that the slow pupil has a greater opportunity for gaining social approval and recognition in an ability grouped classroom than he does in a random grouped classroom. The absence of any significant between-treatment differences in mobility between the second and third year of the study again suggests that adjustments in sociometric status brought about by ability grouping are largely completed by the second year after its introduction.

We were also interested in learning what happened to a pupil's social status in the two grouping systems when his ability level changed. In District A such a change meant a change in classroom assignment, while in District R the pupil would be unaware of the change. First let us consider pupils who received a lower ability level classification in a succeeding year. District A pupils in this classification experienced significantly more upward mobility and less downward mobility than was experienced by the comparable District R pupils. This same trend prevailed in comparisons between the second and third and first and third years although neither attained statistical significance. The movement downward from one ability level classification to another in the ability grouped situation would usually result in a pupil being at the upward extreme of his new ability classification. Thus, the downward movement seemed to increase his relative social status as he was recognized as having come from a higher ability group and was near the top of his new group whereas he had been near the bottom of his previous group. In randomly grouped classrooms it will be recalled that the pupil's ability level classification was unknown to him. He was not subject to room reassignment as a result of obtaining lower ability classification. These data would suggest that any damage to the ability grouped pupil's self-concept that might occur because of his reclassification to a lower ability section might be compensated to some extent by gains in social status he is likely to experience in his new class.

Finally, pupils in District A and District R who moved upward in ability level classification were studied. Two of the three chi-square comparisons made between these pupils yielded statistically significant differences. In each instance, pupils in random grouped classrooms showed more favorable sociometric status changes than comparable pupils in District A classrooms. This difference may well have been caused by the same factor that appeared to operate on ability grouped pupils who moved downward in ability level. That is, when the pupil is moved from the upper extreme of one ability group and placed in the lower extreme of the next higher group his relative status within the classroom may suffer. These findings suggest that although ability grouping might initially provide greater opportunities for social leadership for those of average and low ability, it might also make more difficult the task of maintaining or gaining social recognition if the pupil makes achievement gains sufficient to raise his ability level classification. Loss of social status that is likely to accompany his reassignment to a new classroom might outweigh any gain in status he may experience from being identified as a member of the brighter group.

TABLE 14

CORRELATIONS BETWEEN SOCIOMETRIC STATUS AND OTHER VARIABLES FOR EACH DISTRICT

	Total Sociometric Status	
Other Variables	District R	District A
Mental Ability		
Language Mental Age	.396[a]	.052
Non-Language Mental Age	.239[a]	.002
Academic Achievement	.161[a]	.085
Desirable Attitudes		
Toward School	.234[a]	.121[b]
Toward Teachers	.099[b]	.113
Toward Peers	.355[a]	.311[a]
Concept of Self	.161[a]	.194[a]
Acceptance of Self	.084[b]	.141[b]
Ideal Self	.226[a]	.129[b]
Discrepancy	.013	.079

[a] Significant at .01 level of confidence.
[b] Significant at .05 level of confidence.

Relationships Between Sociometric Status and Other Variables

Correlations between sociometric status and mental ability measured by the California Test of Mental Maturity showed significant relationships for District R pupils on both the language and the non-language scores, while relationships between these variables in District A were close to zero (see Table 14). Thus, mental ability is related to sociometric status in random grouped classes but not in ability grouped classes. These results suggest that the relationship between mental ability and sociometric status reported in previous research may be dependent upon the type of classroom organization. A number of researchers have reported that children with higher levels of intellectual ability tend to receive more sociometric choices than children of lower ability (39, 52, 69, 13, 14, 15). It is, of course, difficult to interpret relationships between mental ability and sociometric choice in ability grouped classrooms because pupils in such classrooms are limited in their choice to other pupils in a restricted ability range.

Correlations were computed between sociometric status and academic achievement, using a composite score based on four of the Sequential Tests of Educational Progress. The results followed the same pattern as the correlations with mental ability, with a significant relationship found in the District R sample and no significant relationship for the District A sample. Correlations between sociometric status and both intelligence and achievement are reduced by the restriction of ability range in the District A classrooms. These data suggest, however, that in ability grouped classrooms the pupil's ability or achievement level is not a factor in determining his sociometric status, while these variables appear to exert some influence on status in random grouped classrooms.

Correlations between sociometric status and the attitudes of pupils toward school, their teacher, and peers were also computed. Low but significant relationships generally existed between desirable attitudes and sociometric status with the highest correlations being found between sociometric status and attitudes toward peers. As the pupil's attitude toward his peers is almost certainly influenced by his sociometric status, it is not surprising that a significant correlation was found between these two variables.

Low but significant relations were also found between the pupil's sociometric status and the various aspects of self concept as measured by Bills' Index of Adjustment and Values. These correlations are very low, however, ranging from .08 to .23 and reflecting very slight relationships between the measures. The presence of these relationships, however, suggests that somewhat larger correlations may exist between the characteristics themselves. This speculation, however, must await the development of more sophisticated measures before it may be tested.

SUMMARY AND CONCLUSIONS

The sociometric data collected for Sample IV pupils during their fourth, fifth, and sixth grade years, and for Sample VI pupils during their sixth-grade year seem to provide considerable evidence in support of the following conclusions relating to sociometric choice in ability grouped and random grouped classes during the intermediate grades:

1. The overall proportions of stars, regulars, and neglectees in District A and R were similar, thus leading to the conclusion that ability grouping does not result in a permanent leadership vacuum in groups of average and slow pupils. Evidence suggests, however, that some time is required for ability grouped average and slow fourth grade pupils to emerge as sociometric stars, thus creating a temporary peer group leadership vacuum at this grade level. No such temporary vacuum was evident for Sample VI during the first year of the study, suggesting that this temporary leadership vacuum if present at all, is of much shorter duration for sixth-grade pupils.

2. Superior students lose some sociometric status when placed in ability-grouped classrooms. This loss is particularly evident in the star classification but is not accompanied by any increase in the neglectee-isolate classification. Data on mobility of sociometric status suggests that superior pupils regain some of this status loss after they have made an adjustment to the ability grouped situation.

3. Average and slow pupils appear to have a far better chance of gaining social recognition in ability grouped classrooms than do comparable pupils in random grouped classrooms. For the slow pupil, ability grouping not only appears to increase the pupil's chances of being classified as a star, but also reduces his chances of being classified as a neglectee-isolate. Sociometric status mobility patterns between District A and R demonstrate a greater number of status changes in ability grouped classrooms. The trend of these changes is generally toward higher sociometric status, further supporting the conclusion that organization of pupils into ability grouped classes brings about marked changes in sociometric status patterns during the first one or two years and generally leads to permanently higher mean sociometric status for average and slow groups.

4. In random grouped classrooms, the pupil's ability level appears to be an important factor in determining his sociometric status. Sociometric status patterns differ considerably for pupils of different ability levels in random grouped classrooms, with pupils of higher ability consistently obtaining more favorable mean scores. The presence of significant correlations obtained when sociometric status was related to ability and achievement for District R pupils further demonstrates the importance of ability to status in the random grouped class.

5. Ability does not appear to be an important factor in determining sociometric status in ability grouped classrooms. In the Utah study sociometric status patterns for pupils of different ability levels in ability grouped classrooms were closely comparable, showing no evidence of real difference. The absence of correlations between ability and status

suggests that these factors are far less important to friendship choices in ability grouped classrooms.

6. The lack of any consistent significant differences between boys and girls suggests that the effects of ability grouping and random grouping on sociometric variables are about the same for both sexes. Thus, it may be concluded that sociometric status and mobility of sociometric status are essentially unrelated to sex in either random grouped or ability grouped classrooms at the intermediate grade levels of the elementary school.

7. Low, but significant correlations exist between sociometric status, pupil attitude, and self-concept variables. These results demonstrate the presence of relationships between the pupil's social acceptance and his perception of himself, his school environment, and his peers.

CHAPTER 8

PUPIL ATTITUDES[10]

THE POSSIBLE EFFECTS of ability grouping on the pupil's attitudes toward school has been a major concern of educators who have been interested in employing ability grouping in their schools. The atmosphere in which the pupil finds himself seems to differ considerably in the ability grouped and random grouped situation as well as being somewhat different at the different levels of the ability grouped situation. Thus, one might well expect to find differences in the manner that pupils react to and perceive the school environments that these treatments provide.

WORKING HYPOTHESES

Three related attitude areas are sampled by the USU School Inventory. These are Attitude Toward Peers, Attitude Toward Teacher, and Attitude Toward School. Working hypotheses relative to these three scales and the other related measures used in the Utah study will now be discussed.

Attitude Toward Peers

1. It was hypothesized that no significant differences in attitude toward peers would occur between superior pupils in the two grouping treatments. Superior District R pupils might be expected to have more favorable attitudes toward peers because of the greater opportunities these pupils have of gaining social recognition in the random grouped classroom. On the other hand, however, superior pupils in District A probably have somewhat more in common with their peers such as similarity of achievement values, better identification with each other, and greater mutual respect that could lead to more favorable attitudes. It was considered likely that these two influences would offset each other thus leading to the working hypothesis.

2. It was hypothesized that District A pupils of average and low ability would have more favorable attitudes than comparable District R pupils. Both of the aforementioned influences appear to operate in the same direction for average and slow pupils. That is, these pupils would have greater opportunities for peer recognition in the ability grouped classroom and also would experience more similarity between their interests and values and those of their peers of comparable ability. Greater differences would be expected to occur between the slow groups, because of the relatively greater improvement of the ability grouped pupil's peer group status at this ability level.

3. It was hypothesized that no within-treatment differences would occur on this score between District A pupils at different ability levels. The factors hypothesized to influence between-treatment differences in attitudes toward peers would tend to reduce differences between the attitudes of District A pupils at different ability levels.

4. Previous research has indicated that in random grouped classrooms peer status is related to ability and it was therefore hypothesized that significant differences favoring the groups of higher ability level would be found in the within-treatment comparisons of District R groups.

5. No significant sex differences in attitudes toward peers were hypothesized.

Attitude Toward the Teacher and the School

Most of the same differences between treatments that seem likely to influence the pupil's attitudes toward teachers appear to operate in the same fashion on his attitudes toward school. Thus, the same differences were hypothesized for these two subscales.

6. It was hypothesized that both superior and slow pupils would have significantly more favorable attitudes toward their teachers and school in ability grouped classrooms than in random grouped classrooms. It would be expected that attitudes toward the teacher would be more favorable in situations where the teacher can give more attention to individual pupil problems and in situations where the curriculum makes it possible for the teacher to lead the pupil to a greater number of satisfying or need-fulfilling experiences and a smaller number of frustrating experiences. Ability grouping has been advocated because of the apparent opportunities for the teacher to provide such situations in more homogeneous classes. Because both grouping situations provide considerable need-fulfillment for the high achieving pupil, it was expected that the largest differences would be found between slow pupils in the two districts.

7. No differences between average pupils in the two grouping treatments were hypothesized because the curriculum in random grouped classrooms is usually well adapted for average pupils, thus suggesting that little real difference between the treatments exists at this ability level.

8. Since the ability grouped classroom would ap-

pear to adapt to the needs of pupils at all three levels with about equal effectiveness, no significant within-treatment differences in attitude toward the teacher were hypothesized for District A.

9. It was hypothesized that superior and average pupils in District R would exhibit similar attitudes toward the teacher while the attitudes of slow pupils would be significantly less favorable. In District R, the average pupil would seem to have the fewest frustrations because the rate and level of teaching is well adapted to his abilities. The high-achieving pupil, however, also experiences considerable need-fulfillment in the random grouped classroom even though the curriculum may be geared at a rate below his capacity. He receives considerable recognition and also can gain satisfaction from working on enrichment materials available.

10. There seems to be some evidence that girls tend to have generally more favorable attitudes toward the total school situation than boys. Thus, it was hypothesized that for both districts significant differences favoring girls would occur in the between sex comparisons on this scale.

PROCEDURE

The USU School Inventory was administered to Sample IV pupils during their sixth grade year; 311 pupils were tested in District A and 617 in District R. As a measure of school attitudes meeting the requirements of this study could not be found, it was decided to develop such a scale. A large number of statements expressing various ideas and feelings about school-related topics were collected from both teachers and pupils. These statements were reviewed and rewritten in a question form that permitted answering "yes", "no", or "?". After pretest and item analysis, 95 items were retained.

The reliability coefficient for the 95 items on the total scale was .94, using the split-half correlation corrected for length with the Spearman-Brown Prophesy Formula. Reliabilities of the subscales were .95, .91, and .82.

Product moment correlation coefficients were also calculated between scores on the three subscales to determine their level of independence. The correlation coefficient between school attitude and peer attitude was .54; between attitude toward teacher and toward peers was .46; and between attitude toward the teacher and toward school was .57. In view of the fairly high reliabilities of the subscales, these intercorrelations suggest that the three subscales of the USU School Inventory though related, are measuring somewhat different facets of the pupil's attitude.

RESULTS

The results of the attitude study were analyzed using a single classification analysis of variance to compare attitude scores of pupils in the two grouping treatments, at different ability levels and of different sex.

Attitude Toward Peers

Comparisons between pupils of similar ability in District A and R on the Attitude Toward Peers scale revealed no differences approaching significance nor did the slight differences found occur consistently in favor of one district or the other. Thus, the two grouping treatments had no significant differential effect upon this attitude as measured by the USU School Inventory. Within-treatment comparisons on the Attitude Toward Peers scale generally supported the working hypotheses. Four of the six differences between pupils of different ability levels in District R were statistically significant with the higher ability group having more favorable attitudes in all cases. Within-treatment comparisons for District A pupils revealed much smaller differences with only one of the six differences reaching statistical significance at the .05 level. These data suggest that attitudes of pupils toward their peers are rather consistently related to ability in random grouped classes. Sex differences were slight on this variable, and no significant sex differences were found.

Attitude Toward the Teacher

The pupil's attitude toward his teacher was considered the most important of the three areas that the USU School Inventory was designed to measure. Of the 95 items on the inventory, 50 were aimed at measuring this aspect of pupil attitude. Both superior groups in District A scored significantly higher on this subscale than comparable District R pupils. Average pupils in the two districts received very similar mean scores with differences falling far short of statistical significance. Slow boys in District A received significantly more favorable scores than comparable District R pupils but slow girls were not significantly different in the two districts. Thus, the working hypotheses were well supported for superior and average pupils and partially supported for slow pupils.

Comparisons between pupils of different ability levels on the Attitude Toward the Teacher score revealed three statistically significant differences in the six comparisons made for District R pupils. These differences all involved superior groups and indicated a tendency for superior pupils in District R to have somewhat more favorable attitudes than either average or slow pupils. In District A no significant within-treatment differences were hypothesized. The results, however, indicated four significant differences on the six comparisons. Three of the four comparisons involving superior students were significant, in all cases favoring the superior group. The fourth significant difference was between average and slow boys; it favored the slow group. Slow girls in District A also showed more favorable attitudes toward the teacher than did average girls although this difference did not reach statistical significance. No significant sex differences emerged on this variable although five of the non-significant differences favored girls as hypothesized.

Attitude Toward School

It was hypothesized that Attitude Toward School scores would favor District A pupils in comparisons at the superior and slow ability levels and that no differences would occur between average pupils in the two treatments. Only one significant difference emerged. This difference was between slow boys in the two districts and favored District A pupils at the .01 level.

No within-treatment differences were expected on Attitude Toward School scores for District A. Three significant differences occurred, all favoring superior pupils as compared with pupils of average or low ability. These differences probably are a reflection of the need-fulfillment provided in both districts for high achieving pupils. Within-treatment comparisons in District R showed a similar pattern with four significant differences emerging, all of which favored the group of higher ability.

It was hypothesized that girls would have significantly more favorable attitudes toward school in both districts than boys of comparable ability. Three of the six comparisons between the sexes yielded statistically significant differences, all in the hypothesized direction. These data provide fairly clear evidence that girls may be expected to show more favorable attitudes toward school than boys, especially at the superior level.

SUMMARY AND CONCLUSIONS

Pupils in Sample IV were administered the USU School Inventory during their sixth grade year. The USU School Inventory contains three scales: Attitude Toward Peers, Attitude Toward the Teacher, and Attitude Toward School.

Comparisons between the two grouping treatments, between pupils of different ability levels within each grouping treatment, and between comparable pupils of the opposite sex were made using single classification analysis of variance. Significant differences between mean scores of various subgroups on the aforementioned measures form the basis for our conclusions.

Results on the Attitude Toward Peers scale revealed no significant between-treatment differences. Four of the six within-treatment differences in District R were significant, favoring the higher ability group in each case. No significant sex differences were found on this scale. These results suggest the following conclusions:

1. The two grouping treatments had no significant effect on attitudes of pupils toward their peers.
2. Attitudes toward peers were found to be consistently related to ability in the random grouped classrooms while no such relationship was found in the ability grouped classrooms, suggesting that ability is more important in determining peer attitudes in the random grouped situation.
3. There was no relationship between sex and attitudes toward peers.

Comparison on the Attitude Toward the Teacher scale showed that superior girls and boys and slow boys in District A received significantly more favorable scores than comparable District A pupils. No significant differences were found between average pupils in the two districts. Within-treatment comparisons between pupils at different ability levels showed three significant differences in District R, all favoring the superior groups. There were four significant differences between District A pupils of different ability levels, three favoring superior groups and one favoring slow boys over average boys. No significant sex differences were found in this attitude area. From these data it may be concluded that:

4. Ability grouping appears to cause more favorable attitudes toward the teacher than random grouping for both superior and slow pupils.
5. Superior pupils in both districts indicated more favorable attitudes toward the teacher than pupils at the lower ability levels.
6. There was some tendency in District A for slow pupils to have more favorable attitudes toward the teacher than average pupils, thus supporting the conclusion that ability grouping has a favorable effect upon the attitudes of slow pupils toward the teacher.
7. There were no differences between the mean attitudes of comparable boys' and girls' groups toward the teacher.

Comparisons between the two districts on the Attitude Toward School scale showed one significant difference favoring slow boys in District A. In both districts there were significant differences between pupils of different ability levels, all indicative of more favorable attitudes for pupils of higher ability. Three of the six sex differences on the Attitude Toward School scale were statistically significant. These differences, all of which favored girls, were largest at the superior level. It may be concluded that:

8. Boys of low ability developed more favorable attitudes toward school in ability grouped classrooms than in random grouped classrooms.
9. Superior pupils had more favorable attitudes toward school in both treatments than pupils of lower ability.
10. Superior girls had more favorable attitudes toward school in both districts than boys of comparable ability.

CHAPTER 9

PUPIL PROBLEMS

THE MAIN purpose of this chapter is to compare the problems reported by pupils in ability grouped and random grouped classes. The instruments employed were the SRA Youth Inventory and the SRA Junior Inventory. These measures are essentially check lists each of which classifies the problems listed into a number of areas. A study of these problem areas on the two inventories revealed several areas in which one of the grouping treatments employed in the Utah study might be expected to lead to greater numbers of pupil problems than the other grouping treatment.

WORKING HYPOTHESES

As no previous research has been reported in this area, working hypotheses were established by weighing the possible psychological effects of ability grouping and random grouping on the various problem areas covered on the SRA Junior Inventory.

The "Looking Ahead" area of the SRA Youth Inventory is concerned primarily with problems related to choosing a career, going to college, and finding a job (94). Many of the problems included in this section are concerned with the pupil's insight into his ability.

1. It was thus hypothesized that District A pupils would indicate fewer problems in this area than District R pupils at all ability levels with greatest differences occurring between superior and average subgroups in the two districts. It appears likely that the ability grouping situation gives pupils at all levels a sounder basis for estimating their ability. Superior and average pupils should be better able than slow pupils to translate this information into answers to the sort of questions raised in this section of the Youth Inventory.

The next problem area to be discussed, "Getting Along with Others", is found in both of the SRA inventories. In the Junior Inventory, the majority of the items relate to social adjustment problems, while in the Youth Inventory, social skills problems are emphasized (95).

2. It was hypothesized, however, that District R pupils would have a greater number of problems on both scales than the comparable District A pupils. Because of the reduced chances of social rejection stemming from ability differences, it would seem that social adjustment problems would be less serious in District A, especially for pupils of low ability. Social skills appear less likely to be influenced by the grouping treatment, but again it appeared probable that the greater opportunities in District A for pupils at the average and slow levels to gain leadership positions and play more active roles in social activities, would tend to result in a lower frequency of social problems.

The "Boy Meets Girl" area of the SRA Youth Inventory is concerned with dating and sex information, and sex mores. It seemed unlikely that the grouping treatments would have any effect on the latter two topics. With regard to dating practices, however, it appeared likely that District A pupils would report fewer problems. Ability grouping at all levels would appear to place pupils in contact with more individuals having similar interests and aspirations. Also, the more favorable social climate present in the average and slow ability grouped sections should tend to reduce dating problems.

3. Thus, it was hypothesized that between-treatment differences in the "Boy Meets Girl" problem area would generally favor pupils in ability grouped classes.

The "Basic Difficulty Key" for the SRA Youth Inventory is made up of those items considered by mental hygiene experts to be indicative of possible personality disturbance (94). Although our working hypotheses on the SRA problem areas generally hypothesized more problems for pupils in random grouped classes, large differences were not expected, and there seemed to be little reason to expect either grouping system to bring about serious personality disturbances.

4. Thus, it was hypothesized that no significant between-treatment differences would occur on the "Basic Difficulty Key".

Three of the problem areas found in the SRA Junior and Youth Inventories were administered only because removing them from the instrument would have exposed subjects to a non-standard administration of the measure that could in turn have changed responses in some of the areas where directional hypotheses were established. No working hypotheses were developed for these areas. Other problem areas such as the "My School" scale, and the "About Myself" scale were measured more adequately by other instruments in the Utah study.[11]

PROCEDURE

Sample

The SRA Junior Inventory, Form S was administered to Sample IV pupils in Districts A and R during their fifth-grade year. The SRA Youth Inventory, Form A was administered to Sample VI pupils during their seventh grade year, Sample VII pupils during their eighth grade year, and Sample VIII pupils during their ninth grade year. It will be noted that for all samples, these data were collected during the second year of the study, after District A pupils had been exposed to about one and one-half years in the ability grouping situation.

Measures

The problem areas in the SRA Junior Inventory include the following: "About Me and My Home", "About Myself", "Getting Along With Other People", and "Things in General". The SRA Junior Inventory, Form S provides a means for the pupil to indicate the seriousness of each problem. In addition to counting the number of problems checked in each of the aforementioned areas, a count of each pupil's "big problems" was made.

The SRA Youth Inventory, Form A was administered to the junior-high-school samples. The Youth Inventory is similar to the Junior Inventory excepting for differences necessary to better fit the problems of junior and senior high school pupils. On the Junior Inventory, Form S the pupil may indicate the seriousness of his problems, while this cannot be indicated on the Youth Inventory, Form A.

The Youth Inventory contains items in eight problem areas. These areas are entitled: "My School", "Looking Ahead", "About Myself", "Getting Along With Others", "My Home and Family", "Boy Meets Girl", "Health", and "Things in General". In addition to these eight problem areas, which contain a total of 298 items, a "Basic Difficulty Key" has been developed which provides another score on the Youth Inventory.

RESULTS

The first score to be discussed, "Looking Ahead", is not included in the Junior Inventory, as questions are concerned primarily with problems related to choosing a career, going to college and finding a job. Typical items in this section include: "How much ability do I actually have?", "Shall I go to college?", "Do I have the ability to do college work?", "What jobs are open to high school graduates?" (95).

Of the 18 between-treatment comparisons of group means in this problem area, there were nine differences that reached or approached statistical significance, all favoring the District A subgroups (see Table 15). As a number of the questions in this area are concerned with ability it is not surprising to note that the pupils in ability grouped classrooms reported fewer problems in this area than pupils in random grouped classrooms. This difference was found most commonly among superior boys and girls and average boys and may reflect a tendency for the District R pupils to have somewhat more doubts as to the level of their ability, thus leading to a larger number of problems.

The "Getting Along With Others" section of the Youth Inventory is concerned with problems related to the pupil's social development with considerable emphasis upon social skills. Typical of the items in this area are: "I wish I could carry on a pleasant conversation", "I want to learn to dance", and "I need to learn how to plan a party". Social skill items make up about two-thirds of the 40 items in this area. Most of the remaining items in this section deal more directly with social adjustment, for example: "I wish I were more popular", and "I am not attractive to the other sex" (94).

The SRA Junior Inventory contains 34 items in this area. In contrast to the Youth Inventory, the great majority of Junior Inventory items relate to social adjustment problems both with adults and peers, with very few pertaining directly to social skills. In view of these differences in the two measures, Sample IV data are not directly comparable to data from Samples VI, VII, and VIII.

Between-treatment comparisons for Sample IV showed District A average pupils, both boys and girls, to have significantly higher mean problem scores in this area than comparable District R pupils. Although not significant, this same trend was found in the mean scores of slow boys and girls.

Results for Samples VI, VII, and VIII who were administered the Youth Inventory, showed a trend for District R pupils to have more problems in this area at all three ability levels. Of the 18 comparisons made at these grade levels three were statistically significant and four approached significance, all in the aforementioned direction. It will be noted in Table 15 that this trend is no more pronounced at one ability level than another.

These data would indicate that at the elementary level, District A pupils listed more social adjustment problems than their District R counterparts, while at the secondary level, District R pupils listed more problems on the Youth Inventory which, it will be recalled, deals more with social skills than social adjustment. The latter trend is consistent enough to indicate a true difference that is general at all three ability levels.

The "Boy Meets Girl" area is not included in the SRA Junior Inventory. Most of the 32 items included in this area of the Youth Inventory are concerned with either dating or sex information and mores with about equal coverage of these two topics. Typical items on dating include: "I don't know how to ask for a date", and "I don't know what to do on a date". Typical items on sex include: "I am embarrassed by any discussion on sex", and "I need more information about sex" (94).

Review of between-treatment differences showed

TABLE 15

BETWEEN-TREATMENT DIFFERENCES ON SRA JUNIOR AND YOUTH INVENTORIES

	Looking Ahead			Getting Along With Others				Boy Meets Girl			Basic Difficulty Key			
Group Sample	VI	VII	VIII	IV	VI	VII	VIII	VI	VII	VIII	IV[c]	VI	VII	VIII
Superior Boys	R[a]	R	NS	NS	R	R	NS	R	NS	R	NS	R	NS	R
t-ratio	AS	AS			AS	1. 96		2. 34		3. 22		AS		AS
Superior Girls	R[b]	NS	R	NS	R	NS	NS	R	NS	R	NS	NS	A	R
t-ratio	2. 08		2. 68		AS			2. 79		2. 07			AS	AS
Average Boys	R	R	R	A	R	NS	NS	R	NS	NS	A	NS	R	NS
t-ratio	2. 84	2. 07	AS	2. 29	AS			2. 79			3. 25		AS	
Average Girls	NS	NS	NS	A	R	NS	NS	R	R	R	A	NS	NS	R
t-ratio				3. 18	AS			2. 02	2. 17	AS	2. 66			2. 70
Slow Boys	NS	NS	R	NS	NS	NS	R	NS	NS	NS	NS	NS	NS	R
t-ratio			2. 67				2. 55							AS
Slow Girls	NS	NS	R	NS	NS	NS	R	NS	NS	R	NS	NS	NS	NS
t-ratio			3. 34				2. 09			2. 03				

[a] Difference between indicated groups approaching significance favoring district indicated (R or A).
[b] Difference between indicated groups significant favoring district indicated and giving "t" value.
[c] Serious problems on Junior Inventory.

a consistent and marked trend for District R pupils to report more problems in this area than comparable pupils in District A (see Table 15). Eight of the 18 between-treatment comparisons were statistically significant and one additional approached significance; all in the aforementioned direction. Differences were generally larger and occurred most often between superior groups and least often between slow groups. Four of the six comparisons between superior pupils were statistically significant as compared with only one significant difference between slow pupils. Although between-treatment differences were hypothesized on this variable, the number of significant differences that emerged was somewhat surprising. Further research will be needed to determine whether the differences found in the Utah study are generalizable to other samples and if so to explore possible reasons why the ability grouped situation tends to be associated with fewer problems being reported on this section of the Youth Inventory.

The Junior Inventory, Form S, administered to Sample IV provided a means for pupils to identify the level of seriousness of their problems. The number of serious problems was obtained separately for pupils taking the Junior Inventory. Two significant between treatment differences were found for Sample IV on this serious problem score, District A average boys and girls listed a greater mean number of serious problems than comparable pupils in District R.

The "Basic Difficulty Key" of the SRA Youth Inventory includes 101 items taken from the Inventory that a panel of seven mental hygiene experts considered to be indicative of a possible basic personality disturbance. For the most part, the remaining items on the Youth Inventory are more a matter of problem recognition. A number of comparisons of "Basic Difficulty" scores with external criteria reported in the test manual would indicate that the scale had some validity as a measure of overall adjustment (94).

Between-treatment results showed that, although only one difference reached statistical significance, six other differences approached significance. All but one of these differences were found in comparisons of superior and average students, and all but one difference indicated a higher basic difficulty score for pupils in District R. It is difficult to be sure on the basis of field study data that these differences came about as a result of the two grouping treatments. These data, however, may lead us to question the claim often made by educators that ability grouping tends to cause pupil maladjustment.

SUMMARY AND CONCLUSIONS

During the second year of the Utah study, Sample IV pupils were administered the SRA Junior Inventory, Form S, and Sample VI, VII, and VIII pupils were administered the SRA Youth Inventory, Form A. The results of these measures were analyzed, using analysis of variance in order to compare the mean number of problems checked by pupils of comparable ability in the two districts. As the SRA Youth Inventory was administered to three different samples at the junior-high-school level, an overall summary of the results on this measure seems desirable.

In reviewing the between-treatment differences on the problem areas of the Youth Inventory reported in this chapter, perhaps the most noteworthy trend is the preponderance of significant differences indicating larger mean problem scores for the District R groups. A total of 32 between-treatment differences reached or approached statistical significance for the junior high school samples. These differences were found in all samples. Of these differences only one indicated more favorable problem scores for District R pupils while the other 31 indicated more favorable problem scores for District A. When the reader considers that these data are based upon pupils in three different junior high schools in each district, and when it is recalled that the samples were comparable on such variables as intelligence and sociometric status, the results would seem to suggest strongly that treatment differences between ability grouping and random grouping in mathematics and science may have a widespread effect upon the number of problems perceived by junior high school students. Whether this difference reflects a greater awareness of problems among District R pupils or reflects a tendency for ability grouping to reduce problems or help pupils deal with them cannot be answered from the data collected in this research.

When the data in Table 15 are summarized by ability level another trend emerges – the tendency for the frequency of significant differences between the districts to be related to ability level. Of the 32 differences reaching or approaching statistical significance on the SRA Youth Inventory, 15 of these differences are found between superior pupils in the two districts, and in 14 of the 15 differences the District R pupils obtained higher (less favorable) scores. Eleven of the 32 differences were between average pupils in the two districts, with all differences indicating higher problem scores for District R groups. These results, reflecting as they do a consistently high number of problems for superior and average District R groups, would suggest the presence of true treatment differences at these ability levels. This trend, although present for slow pupils, is far less pronounced, and less general from sample to sample. Six differences were found between slow pupils in the two districts and all reflected higher problem scores for District R pupils. All six of the between-treatment differences found among slow pupils, however, occurred in Sample VIII, suggesting that they reflect unique characteristics of the sample rather than a generalizable treatment effect. These data support the overall conclusion that ability grouped junior high school pupils in average or superior groups report fewer problems than do comparable pupils in random grouped classes.

Let us now summarize findings in the specific problem areas of the two SRA inventories and see what conclusions may be drawn. These conclusions are better supported for junior high school pupils than for elementary school pupils, because of the availability of comparative data from three junior-high school samples.

Ability grouped junior high school pupils were found to report fewer problems on the "Looking Ahead" scale of the SRA Youth Inventory than comparable pupils in random grouped classes. These differences are most pronounced between superior boys and girls and average boys. It may be concluded that junior high school pupils in ability grouped classes have fewer problems relating to choosing a career, going to college, and finding a job than do comparable pupils in random grouped classes.

The "Getting Along With Others" scale differs somewhat on the two inventories. The results showed two significant between-treatment differences favoring District R pupils on the elementary school data. These results lead to the tentative conclusion that average pupils in the ability grouped classes at the elementary level have more social adjustment problems than comparable random grouped pupils. Three significant differences plus four differences approaching significance were found on the junior high school data, all indicating higher problem scores for District R pupils. At this level, it may be concluded that there is a tendency for random grouped pupils at all ability levels to have a greater number of problems relating to social skills than comparable pupils in ability grouped classrooms.

Comparisons between the treatments on the "Boy Meets Girl" scale showed eight statistically significant differences and one difference approaching significance. These differences all indicated a larger number of problems for District R junior high school pupils. From these data, it may be concluded that there is a consistent tendency for junior high school pupils in random grouped classes to report greater numbers of problems relating to dating, sex information, and sex mores than comparaable pupils in ability grouped classes. These differences were most marked between superior pupils and least pronounced between slow pupils in the two treatments.

Results on the "Basic Difficulty Key" of the Youth Inventory revealed one significant difference favoring District A, and five differences approaching significance in the same direction. These data suggest that if any significant between-treatment difference in pupil adjustment was present, this difference favored the ability grouping system. Because of the small differences found between the districts on the "Basic Difficulty Key", however, it seems safer to conclude that neither treatment is consistently associated with more indications of pupil maladjustment.

CHAPTER 10

SELF-CONCEPT

THE SELF-CONCEPT or phenomenal self as defined in contemporary psychology refers to the individual's perception of himself. That is, expressed in simple terms, the ideas, perceptions, and beliefs that form the image of himself that the individual has created. This picture that he has of himself may be highly realistic in the well adjusted individual, or it can be completely detached from reality as in the case of the psychotic who believes himself to be Napoleon. Regardless of the individual's self-concept, however, he tends to behave in a manner that he considers to be consistent with this concept. The self-concept is considered to be an important factor underlying the individual's behavior by most contemporary personality theorists. Wylie in her review of the self-concept points out that "... all the theories of personality which have been put forth within the last two decades assign importance to a phenomenal and/or nonphenomenal self-concept with cognitive and motivational attributes" (125).

Among theorists who assign an important role to the self-concept are Adler (1), Freud (38) Horney (62), Lecky (71). Maslow (80), Rogers (98) and Snygg and Combs (106). Some personality theorists including Rogers, Snygg and Combs, and Lecky, stress the self-concept as the principal determiner of behavior. Generally, these theorists consider that the need to preserve and enhance the self-concept is a basic human motivation.

Lecky (71) stresses preservation of the self-concept and relates this variable to school behavior. He believes that the child who sees himself as weak, bad, or stupid will behave in ways that will tend to perpetuate this picture and refuse to accept the idea that he is good, strong, or clever. Jersild (65: 114) reflects a similar opinion concerning learning when he points out that "When a person resists learning that may be beneficial to him, he is, in effect, trying to protect or to shield an unhealthy condition. But, more broadly speaking, he is not actually protecting something unhealthy as such; he is trying to safeguard his picture of himself, his self-concept, the illusions concerning himself which he has built and which give him much trouble."

When we view the self-concept as a determiner of behavior and as a picture that the individual will maintain at considerable cost but is under some pressure to change when his behavior is contradictory to it, we can see some definite implications of this theory to the school situation and more specifically to the possible effects of ability grouping. Part of the picture that makes up the pupil's self-concept is concerned with the school. This picture, of course, is highly complex and involves such matters as the individual's estimate of his achievement, his perception of his status relative to the other pupils in the class, his satisfaction with his school role, and his attitudes about study. Each aspect of the pupil's self-perception has subtle shadings and is colored by his value system, the pressures placed upon him by the home and a great many other variables.

The relative impact of the two grouping treatments upon the individual's self-concept appears to be an important factor in their overall evaluation. Ability grouping and random grouping might be expected to influence the individual's self-concept in a number of ways. For example, a pupil who has been unaware of possessing superior ability during the years he has spent in a random grouped system may, upon being assigned to a superior class in an ability grouped system, become more actively aware of his academic status and incorporate this into his self-concept. The pupil of low ability may similarly be successful in a heterogeneous class in maintaining a higher self-perception of his ability than the objective reality of the situation would justify. Upon being placed in an ability grouped system, however, in which this unrealistic self-concept is directly contrary to his placement in a low ability group, he may find it no longer possible to maintain his previous estimate of his own ability and may thus find it necessary to revise downward some aspects of his self-concept.

On the other hand, the increased competition in the ability grouped class may tend to lower the self-concept of the bright pupil while the decreased competition faced by the slow pupil may raise his self-concept. Another circumstance that could lead to revision of the individual's self-concept in an ability grouped situation arises when his initial assignment to a section at one ability level is followed by subsequent reassignment to a lower or higher section.

Thus, it appears that, in comparisons between pupils of the same ability level in the two districts, two major treatment differences would operate that could have an effect upon the self-concept scores of the subjects. These differences are: 1) the awareness of the pupil in District A of the level at which he is placed, and 2) the necessity in District R for the pupil to compete with pupils of all ability levels.

WORKING HYPOTHESES

Our estimate of the impact of these two factors upon the measures of self-concept employed in this research forms the basis for developing our working hypotheses.

The Index of Adjustment and Values, used in this research, yields four scores. These scores are: 1) Concept of Self, 2) Acceptance of Self, 3) Ideal Self, and 4) Discrepancy Score, the latter obtained by subtracting Concept of Self from Ideal of Self.

Applying the aforementioned treatment differences to superior District A pupils, we can assume that they would, in all probability, be aware of their placement in superior classes and this placement should tend to develop higher self-concepts in the individual. On the other hand, placement in the superior section brings with it more intense competition and less of a feeling of being outstanding that the pupil may have experienced prior to the adoption of ability grouping in District A. Because most superior pupils could be expected to be reasonably aware of their ability in either treatment, the latter factor may cause a negative net effect on self-concept.

1. Thus, it was hypothesized that District A superior pupils would develop higher Discrepancy Scores and lower Acceptance of Self scores. In other words, the bright pupil in the ability grouped classroom would be aware of his superiority but would also be aware that the other pupils in his group were equally or perhaps more superior and this awareness would lead him to be less self satisfied that he would be in a random grouped class.

2. It was hypothesized that average pupils in District A would obtain lower scores in Self-Concept and Acceptance of Self and higher Discrepancy Scores because the individual would be more actively aware of his average status and less likely to overestimate his ability than would be the case in District R. Thus, District A average pupils would be somewhat obliged to face reality, but would be in little better competitive position than if they were in a heterogeneous classroom.

It was considered that the same factors hypothesized to influence the self-concept of superior pupils would act somewhat differently upon the slow pupil. Identification with the slow group would tend to lower the pupil's self-concept while membership in the homogeneous class would make him better able to compete and this better competitive position might tend to enhance his self-concept. It was further reasoned, however, that the stigma of being identified with the slow group would out-weigh the advantages gained by the individual's better competitive position.

3. Thus, it was hypothesized that slow pupils in District A would have lower Self-Concept scores, lower Acceptance of Self scores and higher Discrepancy Scores than would slow pupils in District R.

It will be noted that no working hypotheses were developed for the Ideal Self score, as this variable was not expected to change as a result of the grouping situation.

PROCEDURE

The Sample

Pupils in all five samples were administered the IAV. Samples IV and VI were each tested three times over a three year period, while Samples VII, VIII and IX were each tested twice over a two year period. The approximate number of cases for each sample may be found in Table 11.

The Measure

The principal measure of self-concept employed in this research was the Index of Adjustment and Values, developed by Robert E. Bills (11).[12] The Index of Adjustment and Values is currently available in four forms: 1) the elementary school form to be used in grades three, four and five, 2) the junior high school form for grades six, seven and eight, 3) the high school form for grades nine, ten and 11, and 4) the original adult form which is also used in Grade 12.

The high school form contains 37 trait words, the junior high school form contains 35 words, and the elementary school form contains 19. In all of these forms the pupil answers three questions about himself: he rates himself with respect to the trait word (concept-of self), tells how he feels about being this sort of person (acceptance-of-self), and how he would like to be with respect to each trait (ideal-self). In addition to yielding these three scores, a fourth score, Discrepancy, was obtained by subtracting the concept of self score from the ideal self score.[13]

A total of 455 pupils in District R Sample IV completed all three of the administrations of the IAV scheduled for this sample. In order to get some further insight into the relationship between the elementary and junior high school forms of the measure and some estimates of the stability of the IAV scores, the four IAV subscores earned by these pupils on the three administrations of the measure were correlated. The results may be found in Table 16. It will be noted that the correlations between the elementary school IAV administered at the fifth grade level and the junior high school IAV administered at the sixth and seventh grade levels were quite low, ranging from .09 between the fifth and seventh grade Ideal Self scores to .28 between the fifth and sixth grade Concept of Self scores. On the other hand, correlations between the sixth and seventh grade administrations of the junior high school IAV ranged from .41 to .51. These data seem to indicate that the elementary and junior high school levels of the test may be measuring variables that are only roughly comparable. The correlations between the sixth and

TABLE 16

CORRELATIONS BETWEEN IAV SCORES ADMINISTERED IN FIFTH, SIXTH, AND SEVENTH GRADES TO SAMPLE IV, DISTRICT R (N = 455)

	Concept of Self		Acceptance of Self		Ideal Self		Discrepancy Score	
	6th	7th	6th	7th	6th	7th	6th	7th
Fifth Grade	.28	.27	.25	.20	.15	.09	.16	.18
Sixth Grade		.51		.46		.46		.41

seventh grade administrations seem to indicate that the junior high school form of the test has moderately good long term stability.

RESULTS

Sample IV

Sample IV was administered the Bills Index of Adjustment and Values at the fifth, sixth, and seventh grade levels. At the fifth grade level, the elementary form was employed while the junior high school form was administered at the sixth and seventh grade level. The Sample IV data provide us with our best information concerning the relationship at the elementary level between the treatments and the four scores obtained from the Index of Adjustment and Values. All between-treatment data are summarized in Table 17.

Concept of Self Data for Sample IV, during the first year of the project, revealed that the mean scores of all Sample IV groups were very similar, varying less than two points from the lowest to the highest group mean. One significant difference, however, was found between treatments. District R superior pupils scored significantly higher on Concept of Self than pupils of comparable ability in District A. Comparisons between pupils at the average and slow levels in the two districts yielded no significant differences although the District R samples were slightly higher in both instances. Thus, we find that exposure to ability grouping for a little less than one school year seemed to have had little influence upon the self-concept of the pupils involved when these scores were compared with the District R pupils who remained in random grouped classrooms.

Although following somewhat similar patterns, the Concept of Self scores obtained by Sample IV at the sixth-grade level yielded a number of statistically significant differences. District R pupils generally scored higher than District A pupils on this measure. In comparing groups of similar ability, we find District R pupils were significantly higher than their counterparts in District A in four of the six between-treatment comparisons. Only the difference between slow boys in the two districts failed to reach or approach statistical significance.

During the final year of the study, the patterns that had begun to emerge in the sixth grade data became clearer. Mean scores obtained during the seventh grade showed District R pupils to be significantly higher than comparable pupils in District A in four of the six between-treatment comparisons. Within-treatment differences also revealed a consistent pattern in both districts, and for both sexes; showing the superior pupils to have the highest Concept of Self scores, the average pupils the next highest and the slow pupils, the lowest.

In reviewing the Concept of Self data for Sample IV, several important findings seem to emerge. First, the very slight differences existing between the districts during the fifth grade became greater during the sixth grade and still greater at the seventh grade level. This trend could indicate a cumulative effect of the two treatments upon self-concept brought about by an increased awareness of the grouping. It would also suggest that pupils place a higher value upon their achievement status as they mature and move into the junior high school. It is interesting to note that the mean Concept of Self scores of all groups went up between the sixth grade and seventh grade testing, with the exception of District A slow pupils. This exception may lend support to the argument that ability grouping is destructive to the self-concept of the slow pupil as it tends to label or stigmatize him.

Another finding worthy of discussion in reviewing the Sample IV Concept of Self data is the consistently higher scores earned by District R pupils as compared with District A pupils of comparable ability. In every one of the 15 comparisons made between

TABLE 17

SUMMARY OF SIGNIFICANT BETWEEN-TREATMENT DIFFERENCES ON THE IAV

	Sample IV			Sample VI			Sample VII		Sample VIII		Sample IX	
	Grade			Grade			Grade		Grade		Grade	
	5[a]	6	7	7	8	9	8	9	9	10	9	11[a]
Concept of Self												
Superior Boys	R[b]	R	R							A		
Superior Girls		R*		R			R				A*	
Average Boys		R	R						R			
Average Girls		R	R	R	R	R		R			R	
Slow Boys			R*									
Slow Girls		R*	R				R	R				
Acceptance of Self												
Superior Boys	R	R	R		A*	A*						
Superior Girls				R*			R	R	R		A*	
Average Boys	R	R	R							R		
Average Girls				R		R*		R				
Slow Boys	R		R									
Slow Girls				R*	R		R*					
Ideal Self												
Superior Boys												A
Superior Girls				R								
Average Boys		R	R									
Average Girls		R		R				R*				
Slow Boys				R								
Slow Girls							R					
Discrepancy Score [c]												
Superior Boys	A	A	A	A*		A*					A*	
Superior Girls			A*				A*		A		R	
Average Boys	A	A	A						A*			
Average Girls			A		A	A					A*	
Slow Boys											A*	
Slow Girls		A	A			A*						

* Approached significance, (t-value 1.65 to 1.95 inclusive).

[a] As data were not analysed separately by sex for this sample, a significant difference when found is indicated for both boys and girls.

[b] An "R" indicates that the District R subgroup had a significantly higher mean score; "A" indicates a significantly higher mean for the District A group.

[c] High Discrepancy Scores are considered undesirable.

pupils of comparable ability over the three administrations of the IAV, District R pupils obtained the higher mean score and in 11 of these comparisons, this difference was statistically significant. The fact that these differences were slight at the fifth-grade level and fairly large at the seventh-grade level suggests that they cannot be attributed to initial differences in the pupils themselves but are more likely due to the grouping treatments. Treatment effects seem by far the most likely explanation as it will be recalled that the pupils in each district came from several different schools and were found to be reasonably comparable on variables such as sociometric status that could have had an effect upon Concept of Self scores.

Bills considers the Acceptance of Self score to be of major importance in interpreting the IAV. Patterns on the Acceptance of Self scores were generally similar to those found on the Concept of Self scores. District R pupils generally received higher Acceptance of Self scores than District A pupils of comparable ability. Between-treatment comparisons over the three replications revealed that all three of the first year differences were statistically significant beyond the .01 level of confidence in favor of District R pupils. For the last two years, 12 between-treatment comparisons were made. Five of the six differences between boys' subgroups were statistically significant, while none of the differences between girls' subgroups reached significance. It is interesting to speculate on the possible reasons for the greater between-treatment differences between boys' groups.

Although the Discrepancy Score obtained in this study by subtracting the Concept of Self score from the Ideal Self score, is the last score obtained in the IAV sequence, we will treat it at this time because of it's apparent similarity to the Acceptance of Self score. It would appear that the Discrepancy Score measures self-acceptance somewhat less directly and in a different manner than that employed by the Acceptance of Self score itself. For the other scores obtained, the higher scores are indicative of a more favorable self-concept. The reverse is true, however, for the Discrepancy Score as a large score indicates a great discrepancy between the way the individual views himself and the ideal which he would like to attain.

Two analyses were carried out on the Discrepancy Score data. Examination of the results of the first analysis showed that in the majority of cases the low ability pupils had smaller Discrepancy Scores which would indicate better adjustment. As these data seemed to indicate that slow pupils had the best adjustment to their self-concept, and as this part of the Discrepancy Score results did not agree with the Acceptance of Self data, it was decided that a careful recheck should be made of the IAV test papers to see if some spurious element had entered the situation to reduce the size of the Discrepancy Score for low ability pupils. A number of cases were found upon examination of Sample IV scores that suggested that pupils had mechanically checked the responses on the various columns of the IAV. In other words, pupils had followed identical patterns in checking responses on the three columns which usually suggests either a negative attitude concerning the test or a lack of understanding concerning the instructions. Other papers were found in which pupils received a higher score in the Concept of Self column than in the Ideal Self column, thus yielding a negative Discrepancy Score. Bills recommends calculating the Discrepancy Score by taking the difference between the Column 1 and Column 3 entry, regardless of which is larger. This procedure was not followed as it seemed likely that persons whose Discrepancy Scores were negative, i. e., Concept of Self higher than Ideal Self would not be comparable to pupils with equal but positive Discrepancy Scores. Most of the papers yielding negative Discrepancy Scores seemed to indicate a misunderstanding of the instructions, while a few appeared to identify pupils who consciously reported themselves superior to their own ideal self. The apparently spurious papers were most often obtained from pupils of low ability and as such papers yielded a zero or negative Discrepancy Score, they would tend to lower disproportionately the mean Discrepancy Score for low ability pupils. In order to remove these apparently spurious papers, a set of criteria were developed for identifying probably spurious papers and these were removed. After these papers were removed, a second analysis of variance of the Discrepancy Score data was calculated.

This analysis revealed that District R pupils received more desirable Discrepancy Scores than District A pupils of comparable ability in all 15 comparisons made in the three replications of the study. Ten of these 15 differences were statistically significant. Four of the five differences between superior pupils in the two districts reached statistical significance. Significant differences occurred most frequently between superior pupils and least frequently in comparisons between slow pupils. This is probably due in part to the small numbers of cases in the slow groups, as the actual differences between the mean Discrepancy Scores of the slow groups in the two districts were quite large except at the fifth-grade level.

Again it was noted that the between-treatment differences became greater over the three year period of the study. Four of the six between-treatment differences during the third year were statistically significant and the t's for each ability level were higher than the t's for the same ability level during the first year. Thus, a differential effect favoring District R seemed to be operating with respect to changes in the Discrepancy Score. Comparisons of the magnitude of these scores from year to year can best be made between the final two years as the Elementary School IAV administered during the fifth grade is shorter and yields scores that are not closely comparable to those obtained using the Junior High School IAV in the sixth and seventh grades (see Table 17).

The Ideal Self score on the Index of Adjustment and Values calls for the individual to construct an ideal based on the adjectives presented in the scale. In other words, the pupil reports the type of person he would like to be; thus reflecting his aspirations in terms of future self. It was anticipated in this project that the Ideal Self measure would be less likely to be influenced by ability grouping or random grouping than the other scores obtained from the IAV. The mean differences between pupils of comparable ability in the two districts were small during all three years in which the IAV was administered. No significant between-treatment differences were found during the first year. Two significant differences, between the two average groups, were found in the subsequent year. It is interesting to note that for both years two and three, District R average boys obtained significantly higher Ideal Self scores than comparable District A pupils (see Table 17).

Self Concept Changes Accompanying Changes in Ability Level in District A

In view of previous research that indicates most pupils are aware of the ability level to which they are assigned in ability grouped systems, it was hypothesized that reassigment of District A pupils to a lower ability section would be accompanied by changes in the Concept of Self, Acceptance of Self, and Discrepancy Scores on the IAV. In order to test this hypothesis at the elementary school level, pupils in Sample IV who had completed the IAV in both the fifth and sixth grades were classified into three groups: 1) pupils who were assigned to the same ability levels in both grades, 2) pupils who were assigned to a higher ability level in the sixth grade than they had occupied in the fifth grade and 3) pupils who were assigned a lower ability level in the sixth grade than they occupied in the fifth grade. It may be seen in Table 18 that the vast majority of pupils were assigned to the same level during the fifth and sixth grade. A total of 26 pupils for whom complete data were available moved to a higher ability level, while 37 pupils moved to a lower level.

These numbers, although smaller than would be desirable, were considered large enough to justify a comparative analysis of the self-concept data. The fifth grade and sixth grade mean scores on the four IAV subtests earned by these three groups of pupils were compared using analysis of variance and then the sixth grade means were adjusted for fifth grade differences using analysis of covariance to determine whether significant changes in the self-concept scores had taken place.

In reviewing the Concept of Self scores, it is found that groups one and two (pupils assigned to the same or higher level in the sixth grade) obtained significantly higher scores than did group three (pupils assigned to a lower level) on the fifth grade administration of the IAV. This difference, although statistically significant at the .05 level, was small. At the sixth grade level, group three scored significantly lower than the other two groups.

These data indicate that pupils assigned to a lower ability level in the sixth grade were significantly lower in Concept of Self both before and after this reassignment, but that the reassignment itself did not appear to lead to any marked deterioration of the individual's Concept of Self scores. Thus, the lower Concept of Self score may have been a contributing cause to the reassignment of the pupil to a lower ability level or may have resulted from initial misassignment of the pupil to a level where he could not compete. Little evidence is present in the Sample IV data, however, to indicate that assignment of the pupil to a lower ability level brings about a lowering of Concept of Self.

The Acceptance of Self data show fifth grade scores for the three groups to be very similar, with none of the differences between these initial scores reaching statistical significance.

It was tentatively hypothesized that no changes in Ideal Self scores would take place for pupils assigned to the same ability level in subsequent years but that reassignment to a lower ability level would tend to lower Ideal Self scores and assignment to a higher ability level would tend to raise Ideal Self scores. This hypothesis was not supported by the data. Both fifth grade and sixth grade group means on the Ideal Self score were very similar and none of the differences obtained approached statistical significance.

The mean Discrepancy Score, essentially an indirect measure of self acceptance, was expected to go up for group two and go down for group three on the sixth grade testing. Reference to Table 18 shows a trend for group three pupils to have higher Discrepancy Scores on both the fifth and sixth grade administrations of the IAV. The differences at the fifth grade level were very slight, but the sixth grade differences approached statistical significance.

Sample VI

Pupils in Sample VI were administered the junior-high school form of the Index of Adjustment and Values in the seventh, eighth, and ninth grades. The IAV results for this sample provide the best evidence available in this study of the effects of ability grouping upon self-concept at the junior high school level.

Differences between Sample VI pupils on the Concept of Self score on the IAV are summarized in Table 17. For the seventh grade administration of the IAV (when Sample VI pupils had been exposed to one

TABLE 18

SAMPLE IV - CHANGES IN ABILITY LEVEL, FIFTH TO SIXTH GRADE IN DISTRICT A SAMPLE, AND ACCOMPANYING CHANGES IN IAV SCORES

Group	N	Concept of Self Score/Grade		Acceptance of Self Score/Grade		Ideal Self Score/Grade		Discrepancy Score Score/Grade	
		5th	6th	5th	6th	5th	6th	5th	6th
1. Pupils assigned same ability level, fifth and sixth grades	248	44.25	85.05	46.77	89.03	49.72	95.95	5.93	11.93
2. Pupils assigned higher ability level in sixth grade	26	44.04	84.15	45.73	89.04	50.11	95.04	6.08	12.43
3. Pupils assigned lower ability level in sixth grade	37	42.00	81.43	45.86	85.70	49.43	96.67	7.24	15.32

year of ability grouping or random grouping at the sixth grade level and most of a year at the junior high school level) two significant between-treatment differences emerged. District R superior girls were found to be significantly higher in Concept of Self than District A superior girls, and District A average girls were significantly higher than girls of comparable ability in District A. At the eighth and ninth grade testing, District R superior girls again made higher scores than District A superior girls, but these differences did not reach statistical significance. In the case of average girls, however, the District R sample was significantly higher than comparable District A girls in all three administrations of the junior high school Index of Adjustment and Values. Of the nine comparisons between girls of comparable ability in Districts A and R, District R girls received higher Concept of Self scores in every comparison. In contrast, there were no significant between-treatment differences for comparable boys groups in District A versus District R. These data suggest that differences in the two grouping treatments had a somewhat greater effect upon girls than upon boys with regard to Concept of Self scores.

Scores on the Acceptance of Self section of the IAV again seem to reflect the greater impact of the grouping treatments on girls' subgroups. No significant between-treatment differences were found in the nine comparisons between District R and District A boys of comparable ability. Differences between comparable girls' groups in the two districts, revealed that of the nine comparisons made, all favored District R pupils. Two of these differences were statistically significant and three more approached significance.

It will be recalled that Discrepancy Scores in Sample IV were reanalyzed after removing scores that, upon examination of the answer sheets, appeared to be spurious. Such answer sheets were removed from the Sample VI data prior to the start of the analysis so that scores for these cases were not included in any of the IAV subtests for Sample VI.

For Sample IV, rather large and consistent Discrepancy Score differences favoring District R were found in the between-treatment analysis. The between-treatment differences in Sample VI generally followed the same pattern but were much smaller with only two differences, both between average girls in the two districts, being statistically significant.

Between-treatment differences reaching or approaching significance increased during the final year of the study. Even in this final year, however, these differences were considerably less marked than were found in Sample IV, suggesting that the grouping treatment at the junior high school level had somewhat less of an effect upon Discrepancy Scores than the more extensive differences that were present at the elementary school level.

Results for Sample VI on the IAV Ideal Self score are also summarized in Table 17. Of the nine between treatment comparisons for boys, only one, favoring District R slow boys, was statistically significant. A somewhat more definite pattern emerged, however, in the between-treatment differences for girls. In seven of the nine comparisons, the District R girls scored higher, with two of these differences reaching statistical significance.

Self-Concept Changes Accompanying Ability Level Changes in District A

In order to determine whether the four self-concept variables were affected by changes in the pupil's ability level in ability grouped classes, Sample VI pupils were classified into six groups: boys and girls assigned to the same ability level in the sixth and seventh grades, boys and girls assigned to a higher ability level in the seventh grade, and boys and girls assigned a lower ability level in the seventh grade. Analysis of covariance was carried out for each of the four self-concept variables to determine whether the final adjusted mean scores on these variables were significantly different for like-sex pupils whose ability classification remained the same, went up, or went down. The results of this analysis may be found in Table 19. As was the case with Sample IV, the majority of pupils were assigned to the same ability levels in the sixth and seventh grades. A sufficient number of both boys and girls, however, were assigned either higher or lower ability levels to permit an analysis of these data.

In comparing the Concept of Self scores for the sixth grade administration of the IAV, it was found that boys assigned the same level (group 1) were significantly higher than boys who were subsequently assigned to a higher level or a lower level. On the seventh grade administration, however, these initial differences had largely disappeared.

In contrast, the three girls' groups were closely comparable on the sixth grade Concept of Self score but differences had emerged by the seventh-grade administration. Comparison of the adjusted seventh-grade means of girls' groups on the Concept of Self score revealed that girls assigned the same ability level in both the sixth and seventh grades (group 2) had higher mean scores than both the girls assigned to a higher level (group 4) and the girls assigned to a lower level in the seventh grade (group 6). This latter difference was significant at the .01 level, and would suggest that in an ability grouped situation, girls find remaining in the same ability level less threatening to self-concept than either moving to a higher level or a lower level, with movement to a lower level having the least favorable effect upon their scores.

Scores on Acceptance of Self showed that group 1 boys again had a somewhat higher mean score than either group 3 or group 5 boys, but these differences were not statistically significant. At the seventh-

TABLE 19

SAMPLE VI - CHANGES IN ABILITY LEVEL, SIXTH TO SEVENTH GRADE IN THE DISTRICT A SAMPLE, AND ACCOMPANYING CHANGES IN IAV SCORES

		Concept of Self			Acceptance of Self			Ideal Self			Discrepancy Score		
		Mean Score			Mean Score			Mean Score			Mean Score		
		Grade			Grade			Grade			Grade		
Group	N	6th	7th	Sig.[a] Dif.	6th	7th	Sig. Dif.	6th	7th	Sig. Dif.	6th	7th	Sig. Dif.
1. Boys assigned same ability level, sixth and seventh grades	154	89.90	88.04		93.94	89.86		99.57	103.31	gr. 3 3.11	14.75	11.92	
2. Girls assigned same ability level, sixth and seventh grades	157	88.19	89.33	gr. 6[b] 2.63	88.75	90.10	gr. 6 3.40	99.50	101.20	gr. 6 2.77	14.30	12.05	
3. Boys assigned higher level in seventh grade	40	83.77	85.60		88.07	88.45		94.80	95.35		13.02	10.60	
4. Girls assigned higher levels in seventh grade	50	84.08	87.16		86.68	87.88	gr. 6 2.01	97.12	99.50		13.18	12.60	
5. Boys assigned lower level in seventh grade	31	85.41	87.55		90.81	89.74		95.74	97.87		11.64	10.32	
6. Girls assigned lower level in seventh grade	31	84.00	85.35		83.90	82.52		99.32	97.32		14.87	12.42	

[a] Significant differences in final means adjusted for differences in initial means, using covariance analysis.

[b] Indicates that this group is significantly lower than group in left hand column; number indicates size of t.

grade testing, we find that the three boys' groups had become more closely comparable on this score. A comparison of the adjusted seventh grade means for the girls' groups showed both group 2 and group 4 to have significantly higher adjusted Acceptance of Self means than group 6. Thus, it appears that for girls, being assigned to a higher ability level did not result in a significant increase in Acceptance of Self scores, but assignment to a lower level brought about a relative decline in these scores. These data generally support the Sample IV findings.

Initial Ideal Self scores showed the boys who were to remain at the same level to be significantly higher than those who were to be assigned to a higher or lower ability level during the subsequent year. The group 1 boys were significantly higher than group 3 on the final adjusted Ideal Self mean. Group 3 and group 5 boys were not significantly different on the final adjusted mean. Comparisons between the three girls' groups showed them to be quite similar in their sixth grade mean scores with somewhat larger differences occurring in the seventh grade means. The final adjusted mean score for group 2 was significantly higher than that for group 6 on the Ideal Self score. It is interesting to note that although the final adjusted means for group 4 and group 6 girls were not significantly different, the girls who were assigned a higher ability level experienced an increase of about two points between the sixth and seventh grade tests while the girls who were assigned a lower level in the seventh grade experienced a decrease of two points on Ideal Self. These data suggest that for the girls' groups, assignment to a lower ability level tends to lower the individual's Ideal Self score.

No significant differences in Discrepancy Scores were found among pupils who moved up, down, or remained the same in ability level classification between the sixth and seventh grades. All groups had slightly lower Discrepancy Scores at the seventh grade level.

Sample VII

In reviewing the between-treatment differences in Concept of Self for Sample VII, we find District R superior girls have significantly higher Concept of Self scores than District A superior girls for the eighth grade testing (see Table 17). Other between-treatment differences were not statistically significant although three differences approached statistical significance, all favoring District R girls. When one considers that only one between-treatment difference reached statistical significance at the .05 level out of the 12 comparisons made, it seems possible that this single significant difference is due to chance. The fact, however, that the three differences which approached significance also favored District R, suggests that if any self-concept difference did exist, it favored the random grouped sample.

Between-treatment differences on the Acceptance of Self score for Sample VII revealed three significant differences, all favoring District R girls as compared with District A girls. For both administrations of the IAV, District R superior girls were significantly higher than District A superior girls. During the 1960-61 year, District R average girls were significantly higher than District A average girls. The relatively small number of slow girls in the two samples led to a considerably larger standard error of the difference between means for these groups, thus requiring greater differences to reach statistical significance. A rather steady trend appeared to be present in all samples for District R slow girls to score higher on Acceptance of Self than District A slow girls. These results suggest that a true difference is present although this difference reaches statistical significance in only about half of the replications.

Comparisons between the mean Discrepancy Scores of subgroups in the two districts revealed no statistically significant between-treatment differences. Between-treatment comparisons for Sample VII on the Ideal Self score revealed only one statistically significant difference favoring District R slow girls.

Sample VIII

Only two significant between-treatment differences occurred in the Sample VIII Concept of Self scores. During the ninth grade, District R average boys scored significantly higher than District A average boys. During the tenth grade, District A superior boys scored significantly higher than boys of comparable ability in District R.

Comparison of Acceptance of Self means revealed only one statistically significant between-treatment difference for each of the two years in which the IAV was administered to this sample. District R superior girls were significantly higher during the ninth-grade year, and District R average boys were significantly higher during the tenth grade year.

Only one significant between-treatment difference on Discrepancy Score means was obtained for this sample. District A superior girls had a higher mean than comparable District R girls. No significant between-treatment differences were found on the Ideal Self scores. The presence of only one significant difference in the tenth grade data suggests that at the high school level the somewhat changed situation brought about by the wide range of electives available in both districts and less distinct ability groupings in District A, probably reduced the importance of the grouping treatments in terms of their influence upon pupil's self-concept.

Sample IX

The Index of Adjustment and Values was administered to Sample IX pupils during their ninth and eleventh grade years. Analysis procedures for this sample followed the same pattern employed with previous samples excepting that the groups were not divided by sex for the analysis of the eleventh grade data.

Between-treatment comparisons of Concept of Self scores for Sample IX are found in Table 17. Only one significant between-treatment difference was found, ninth grade average District R girls scoring higher than comparable girls in District A. No statistically significant differences were found in the between-treatment comparisons of Sample IX pupils on the Acceptance of Self score. Between-treatment comparisons of the Discrepancy Score showed a consistent tendency for District A pupils to obtain higher (less favorable) scores than comparable pupils in District R. Only the difference between superior girls reached statistical significance but three other differences approached significance. Only one significant between-treatment difference was found in Ideal Self; superior pupils in District A scoring higher than those in District R.

In reviewing the data for Sample IX, we find that the ninth grade results are generally in line with those found in other samples. The almost complete lack of significant differences between IAV scores of different subgroups at the eleventh grade level is probably caused by the small number of cases and the fact that the actual differences between the ability grouping and random treatments is much smaller at this level than in the junior high school and elementary grades. Similar results were found for the Sample VIII high school data.

SUMMARY AND CONCLUSIONS

All five samples in the Utah study were administered the IAV. Samples IV and VI were tested three times over a three year period, while Samples VII, VIII, and IX were tested twice over a two year period.

Between-treatment differences on the four scores obtained from Bills' Index of Adjustment and Values are summarized in Table 17. For each of the four IAV variables, 66 between-treatment comparisons were made, 22 at each ability level. The largest number of between-treatment comparisons, 18, were made for Sample VI pupils. Fifteen comparisons were made for Sample IV, 12 for Samples VII and VIII, and nine for Sample IX. Data collected from the sixth through the ninth grade levels probably provide the most meaningful results on self-concept. Differences between fifth grade pupils were based on the elementary school form of the IAV which is shorter and somewhat less reliable than the junior high school form. Differences between pupils in the high school are less meaningful because of the less clear-cut distinction between the two grouping treatments at that level. Except for data collected on fifth grade pupils in Sample IV, and eleventh grade pupils in Sample IX, boys and girls were placed in separate groups for the IAV analyses.

Of the 66 between-treatment comparisons on the Concept of Self score, 23 reached or approached statistical significance. All but two of these 23 differences were in a direction favoring District R pupils. Eight of these differences were between superior pupils in the two districts, ten between average pupils, and five between slow pupils. It is interesting to note in the Concept of Self analysis that 15 of the 23 differences reaching or approaching statistical significance were between girls' groups while only seven of these differences were between boys' groups. One of the significant differences, for Sample IV, was based on groups not divided by sex. The extent and consistency of the Concept of Self data seem to justify the conclusion that pupils in random grouped classrooms tended to obtain more favorable Concept of Self scores than comparable pupils in ability grouped classrooms, at all ability levels. The data also appear to support the conclusion that the two grouping treatments had a somewhat greater effect upon girls than boys. These data may reflect the greater preoccupation of girls with achievement and academic success. The results generally suggest that placement in an ability group tends to adversely affect girls' self-concept. This adverse effect is of least importance for pupils at the lower ability level.

Of the 66 between-treatment comparisons of IAV Acceptance of Self mean scores, 15 were statistically significant and seven others approached significance. These results reflected essentially the same trend found in analysis of the Concept of Self data. Six of the significant differences were between superior pupils, six between average pupils, and three between slow pupils in the two districts. All 15 of these significant between-treatment differences favored District R as did four of the seven differences that approached significance. It will be noted that the significant differences occurred between girls' groups and boys' groups with about equal frequency. When we consider all results relative to acceptance of self obtained in this study, a consistent pattern favoring District R pupils emerges. This pattern is most pronounced for superior and average pupils but is present to a lesser degree for slow pupils. Based on these data it must be concluded that random grouping is consistently related to higher self acceptance for pupils at all ability levels and over most of the grade levels covered in this project.

The overall summary of between-treatment comparisons on the IAV Discrepancy Score shows 14 of the 66 differences to be statistically significant. Six of these differences were between superior pupils in the two districts. Six were between average pupils and two were between the slow pupils. All but one of the 14 significant differences showed higher (less favorable) *Discrepancy Score* means for District A subgroups. In addition to these significant differences, eight differences approached statistical significance, and all of these showed higher mean *Discrepancy Scores* for District A groups. Significant differences occurred with about equal frequency for boys' and girls' groups. It may be concluded that the two grouping treatments had a differential effect upon the pupils' *Discrepancy Scores* with ability grouping generally associated with less favorable scores. This pattern is most pronounced in the differences between the mean scores of superior pupils in the two treatments, closely followed by average pupils. The differences between slow pupils are substantially smaller.

Of the 66 between-treatment comparisons of mean *Ideal Self* scores on the *Index of Adjustment and Values*, eight reached statistical significance. Seven of these eight differences favored District R pupils, with two occurring between superior groups, four between average groups and two between slow groups. It will be noted in Table 17 that the few significant differences on the *Ideal Self* scores are generally scattered. Only in the comparisons between average girls do we find significant differences occurring in more than one sample. From these data we may conclude that a slight trend favoring District R pupils is in evidence on the *Ideal Self* scores. This trend, producing relatively few significant results, suggests that the grouping system has little or no influence upon the *Ideal Self* score.

Self-Concept Changes Accompanying Ability Level Changes in District A

Pupils in District A, Sample IV who changed ability levels between the fifth and sixth grade administrations of the *IAV* and pupils in Sample VI who changed ability levels between the sixth and seventh-grade administrations of the *IAV* were compared to determine whether changes in ability classification would be accompanied by changes in the self-concept variables. The results of these analyses may be found in Tables 18 and 19. The reader is reminded that the Sample VI data are generally more meaningful because sufficient cases were available to divide the subgroups by sex, and because the junior high school form of the *IAV* was administered at both sixth and seventh grade levels, yielding more comparable data than were obtained for Sample IV. From these analyses, it may be concluded that:

1. Placement of a pupil in a lower ability classification appears to bring about lower scores in *Concept of Self*, *Acceptance of Self*, and *Ideal Self*.

2. The *IAV* scores of girls assigned to a lower ability group appear to be more drastically affected than are the scores of boys under the same circumstances. This may reflect the tendency for girls to place a higher value on academic achievement and status.

3. Pupils destined to be reassigned to a lower ability classification show lower scores in *Concept of Self* and *Acceptance of Self* prior to reassignment. This lower initial score may indicate that misassignment of a pupil to a higher group than his ability would justify brings about a lowering of his *Concept of Self* and *Acceptance of Self* because of his inability to perform satisfactorily in his assigned group.

4. Remaining in the same ability classification in an ability grouping system appears to be less threatening to the pupil's self-concept than reclassification to either a higher or lower section. Reclassification to a higher section may lead to a slight loss in the pupil's self-concept because he is probably near the bottom of the ability range in his new group. This self-concept loss may be compensated to some extent by the pupil's identification with the higher ability classification. The effect upon pupils assigned to a lower ability group, however, appears to be much greater. These pupils probably lose self-concept because of their initial misassignment and then their self-concept is further lowered, especially for girls, when they are confronted with the stigma of identification with a lower ability group.

Whether the less favorable scores on the self-concept variables that occurred rather consistently in the ability grouped samples are considered "good" or "bad" depends on one's point of view. When considered in the light of psychological research in related areas such as aspiration level, however, the author must conclude that the apparent effects of ability grouping upon the self-concept variables that emerged from the Utah study are probably harmful to the development of at least some of the pupils who are educated under such a system of grouping.

CHAPTER II

PERSONALITY

ALTHOUGH many of the arguments heard among educators on the merits and demerits of ability grouping imply that this practice can have an effect upon the personality development of pupils involved, very little direct research evidence comparing the personality characteristics of pupils in ability and random grouping systems has been published up to the time of this writing. Subjective appraisal of the differences that appear to be present in ability grouped and random grouped classes along with the consideration of the limited evidence on the relative merits of these systems suggest possible personality differences that might occur as a result of differences in grouping treatment.

It might seem a simple matter to identify a test measuring the characteristics that we hypothesize are affected by the different grouping treatments, administer the test to pupils in the two treatments and determine whether differences between the scores are statistically significant. This is essentially the approach used in this project, but the reader should be aware of the many limitations attending this approach in the area of personality measurement. First, the greatest differences between the grouping systems studied in this research were at the elementary school level where the self-contained classroom system resulted in placing District A pupils in a single ability level classification for the entire school day. Unfortunately, however, at this level very few personality measures are available, and those available can be regarded at best as very crude instruments. At the junior high school level, a greater number of personality measures are available but it will be recalled that, as District A employed grouping only in mathematics and science at this level, the treatment differences were much less marked than at the elementary level and consequently would be expected to bring about smaller differences in personality.

Another practical problem arising in the study of personality differences between pupils in the two grouping treatments is the question of how much change in personality we might expect to take place as a result of changing a fairly small portion of the individual's total experience. Let us suppose, for example, that one grouping treatment did tend more frequently than the other to develop inferiority feelings. It would seem a strong possibility that the many other experiences in the pupil's day that may structure personality and that cannot adequately be controlled could mask any such personality differences that resulted from the grouping treatment.

A related question that must be faced is whether or not personality changes that might be brought about by the two grouping treatments could be expected to become apparent in a period of one or two years? The lack of any basis for answering this question, of course, leaves us with the possibility that the generally negative results obtained could reflect an inadequate exposure of the subjects to the treatment differences.

Another factor to be considered is the generally low level of our knowledge concerning the nature of personality and its measurement. Evidence of validity is rather scant for most personality measures, and even when validity data are given these data usually are based upon correlations with clinical judgement or with other criteria that themselves are of questionable validity.

Some psychologists, though condemning the personality inventory, place great confidence in projective tests as a source of personality information. This project employed both group administered personality inventories and individual and group administered projective measures. The unbiased researcher, however, must conclude that neither the personality inventory nor the projective measure at its current state of development provides very solid evidence. Thus, the personality findings presented in this section should be considered at best tentative.

WORKING HYPOTHESES

Working hypotheses were established and several personality measures were studied in order to identify variables that appeared to measure the personality characteristics that might be influenced by the two grouping treatments. Variables from four personality inventories were employed: the California Test of Personality (CTP), Inventory of Factors GAMIN (GAMIN), Objective Analytic Personality Tests (OA), and the California Psychological Inventory (CPI). A projective measure employing cards from the Thematic Apperception Test (TAT) and the Michigan Picture Test (MPT) was also used to measure three variables that seemed particularly important.

Previous research evidence provided little information or help in establishing working hypotheses are therefore quite tentative.

1. For the first variable, Feeling of Belonging, (CTP) it was hypothesized that pupils at both the superior and slow ability levels would obtain more

favorable scores in ability grouped classrooms. It would appear that these pupils would experience a greater feeling of belonging in the less heterogeneous District A groups. Slow pupils, especially, would seem more likely to feel secure in a group of other slow pupils. A gradual deterioration of this characteristic for pupils in slow groups might occur, however, outside of the classroom. Ability grouping may cause both bright and slow children to sense that they are "different". For the slow child this difference carries with it the label of inferiority that is implied by his membership in the low ability class. Thus, the very act of separating pupils into groups on the basis of ability might make them actively aware of differences which could in turn lower their "feeling of belonging".

2. It was hypothesized that low ability pupils would display significantly higher scores on Withdrawing Tendencies (CTP) and Anti-Social Tendencies (CTP) in the ability grouped situation. Both seemed likely to be reactions to frustration. Evidence obtained in Drews' study (31) that showed much higher levels of class participation for slow ability grouped students would suggest that such students would be less likely to experience frustration in ability grouped classes. Studies by Jackson (63) and Marsh (61) also reported fewer anti-social behavior patterns in ability grouped situations. Mathias' work (81) which was discussed in the review of literature, would also suggest that less aggression would occur in ability grouped classes.

3. Our working hypothesis with regard to average pupils on the aforementioned CTP variables and to superior pupils on the latter two was that no significant between-treatment differences would occur.

4. It was hypothesized that scores on ascendancy in social situations, (GAMIN) would be higher for superior pupils in District R because of the greater opportunities for these pupils to obtain positions of social leadership in random grouped classrooms. Slow pupils, on the other hand, would appear to have a greater opportunity for social leadership in the ability grouped classroom.

5. Thus, it was hypothesized that slow pupils in District A would obtain higher social ascendancy scores than comparable pupils in District A.

6. It was hypothesized that District A slow pupils would receive less favorable scores on a measure of inferiority feelings (GAMIN) than comparable pupils in District R.

7. It was also hypothesized that superior District A pupils would receive less favorable "inferiority" scores than District R pupils, although this difference might be expected to be less pronounced than the difference between the slow groups. For the slow child, many educators have suggested that the labeling that accompanies ability grouping would tend to lower self-confidence and develop inferiority feelings. Some rather limited research evidence supports this contention (72). The superior pupil might also be expected to have greater self-confidence in the random grouped class because of the smaller number of pupils in the class able to compete with him academically and the greater chance he has of obtaining a position of social recognition within the peer group.

8. It was hypothesized that superior pupils in District R would obtain higher scores on Competent Assertiveness (OA) than superior pupils in District A. Cattell, (25) in discussing this factor suggests that it is probably quite similar to achievement motivation as studied by McClelland. The random grouping situation would seem more likely to develop the self perceptions underlying this factor in superior pupils than would the ability grouping situation.

Anxiety to Achieve (OA) although corresponding to some degree to the clinical concept of anxiety, was found by Cattell (25) to be loaded also with verbal ability, and scholastic achievement (holding intelligence constant). It was hypothesized that superior pupils in District A and slow pupils in District R would be higher on this variable. Both of these groups are subjected to considerable pressure to achieve and thus might be expected to display high levels of achievement-related anxiety.

Working hypotheses for the California Psychological Inventory were developed for three of the four classes of variables measured by this instrument. Class I includes six measures in the broad area of poise, ascendancy and self-assurance.

9. For the Class I variables, it was hypothesized that slow pupils would score higher in the ability grouped classrooms, superior pupils would score higher in the random grouped classrooms and no differences would be found for average pupils. The ability grouping situation seems to provide the low ability pupil considerably more opportunity for developing some of the Class I characteristics. Dominance, for example, seems to have a much greater chance of emerging in a situation where the slow pupil is participating with peers of similar ability. Differences in grouping treatments might also be expected to have some influence on the slow pupils Capacity for Status, Self Acceptance, and Sociability and should contribute directly to his Social Presence. This is supported by Drews' research (31) which indicated that the slow pupil has more opportunity to experience successful social interaction in the ability grouped situation.

Ability grouping, on the other hand, appears to create an unfavorable climate for superior pupils to develop the Class I personality traits. When removed from their favored position in the random grouped classroom, it would appear that the superior pupil's opportunities to develop Dominance, Social Presence, and Sense of Well Being, would be reduced. Sociability might similarly be reduced for the superior pupil in the ability grouped situation as it would seem likely that he would be less secure, less able to dominate the social situation and less able to gain social experience in the more homogeneous class.

Three of the scales on the CPI make up Class III – measures of achievement potential and intellectual efficiency.

10. For two of these measures, Achievement via conformance (Ac) and Intellectual efficiency (Ie) it was hypothesized that superior pupils in ability grouped classrooms would receive higher scores. Ability grouping with acceleration, where pupils are exposed to a uniform program at a rate close to their capacity would seem to develop these two characteristics while discouraging the third Class III variable, Achievement via independence (Ai).

11. Thus, it is hypothesized that on the latter variable superior pupils in random grouped classrooms would achieve more. Slow pupils in District A were also hypothesized to obtain higher Ac scores than their District R counterparts. No other between-treatment differences were hypothesized for the Class III variables.

Class IV measures are described by the test author as those dealing with intellectual and interest modes. A working hypothesis was developed for only one of the three Class IV measures, Flexibility.

12. It was hypothesized that the District R classroom environment with enrichment would tend to develop flexibility to a greater degree at all ability levels than would the ability grouped classroom with acceleration. No working hypotheses were developed for the six measures of socialization, maturity and responsibility which make up the Class II variables on the CPI.

Projective Data

In addition to obtaining data in the Utah study on self-report personality inventories, two aspects of the study were concerned with the measurement of specific personality variables using a projective approach.[14] The first of these studies employed an individually administered measure consisting of eight pictures selected from the Thematic Apperception Test (TAT) (87) and the Michigan Picture Test (MPT) (83). This study was concerned primarily with comparisons between scores of aggression, depression, and inferiority feelings obtained from slow fifth grade pupils in Districts A and R, and will be referred to as the fifth grade study. The second study adapted substantially the same projective measure for group administration. This study involved sixth grade pupils at all three ability levels.

Mathias (81), found that in heterogeneous classrooms, superior and slow learners showed more signs of aggression and depression than did average pupils. Although his study contained no comparable data on ability grouped children, it had implications for ability grouping as he concluded that superior and slow learning children showed more aggression and depression than the average children because the educational program was not designed for them. This would infer that in an ability grouped situation where the program was more closely adapted to the pupil's ability level, less aggression and depression would be found.

Luchins and Luchins (72) reported that children in ability grouped classes felt stigmatized by being placed in the low ability group. The inference here is that feelings of inferiority were increased by ability grouping. However, this study lacked a comparable group from heterogeneous or random grouped classrooms so it is possible that similar feelings of inferiority may have developed in low ability pupils in random grouped classes, although perhaps for different reasons.

The following working hypotheses refer to the fifth and/or sixth grade studies as indicated. It was hypothesized that:

1. More evidence of aggression would be found in slow learning pupils who were randomly grouped than in slow learning pupils who were placed in ability groups (fifth and sixth grade studies).
2. There would be more evidence of depression in slow learning pupils who were randomly grouped than in slow learning pupils who were ability grouped (fifth and sixth grade studies).
3. There would be more evidence of inferiority feelings in slow learning pupils who were ability grouped than would be found in slow learning pupils who were randomly grouped (fifth and sixth grade studies).
4. Superior pupils in random grouped classrooms would obtain higher (less favorable) mean scores on aggression and depression than similar pupils in ability grouped classrooms (sixth grade study).

PROCEDURE

Parts of the California Test of Personality (CTP) elementary level, were administered to Sample IV pupils during their fifth grade year. This test is a fairly direct self-report type of personality inventory. Two subtests, Feeling of Belonging and Withdrawing Tendencies from the Personal Adjustment areas; and one subtest, Antisocial Tendencies from the Social Adjustment area were used to test hypotheses in this study. Reliability coefficients for the subtests of the California Test of Personality, elementary level, that were employed in this research ranged from .77 to .83.

As is the case with most direct self-report measures, the items themselves present some evidence of content validity. Such measures, however, are dependent upon the accuracy of the subject's insight and his willingness to provide honest answers. In the Utah study this measure was administered in a permissive atmosphere and pupils were given no motive for falsifying or faking their responses. In the more important area of concurrent validity little evidence is available on this measure.

The Guilford-Martin Inventory of Factors GAMIN provides scores on five personality variables developed through factor analyses by Guilford and his associates (84). Two of the variables on this measure purport to test characteristics that educators have often suggested were influenced by ability grouping. These are Factor A, which measures ascendancy in social situations as opposed to submissiveness and Factor I, which measures lack of inferiority feelings. Items measuring these two factors were administered to Sample VI pupils during their seventh-grade year, Sample VII pupils during their eighth-grade year, and Sample VIII pupils during their ninth-grade year. In each case, District A pupils had been exposed to the ability grouping treatment for a little more than one and one-half school years. Although this measure is usually not administered below grade nine, the reading level and content of the items appear generally appropriate for pupils as low as grade seven. Factor A contains 38 items and the split-half reliability of this measure as reported in the test manual is .88. Factor I contains 49 items and has a split-half reliability of .89 (55).

At the time that the test battery for this research project was being assembled, the Objective-Analytic Personality Test Batteries by R. B. Cattell and his associates had been available to psychologists for about three years. These measures, although in an experimental form, appeared to offer certain advantages over both the group administered personality inventories and the projective techniques. The factors forming the basis of this battery emerged from a series of factor analysis studies by Cattell and a number of other researchers. This research resulted in the identification of 18 personality factors, the first 12 of which Cattell considers sufficiently invariant and clearly proved by research to be used with confidence (25). Fewer factors have been identified for children, and the factors that have emerged for children are based on fewer studies than those for adults. Cattell has provided a number of individual and group tests that can be used to measure each personality factor in his battery. Children's tests often differ from the adult tests in level of sophistication and in some cases in factor loadings.

Two factors, taken from the 12 Cattell considers best established, were selected for use in this research. The factors were UI-16 Competent Assertiveness, and UI-24, Anxiety to Achieve. Cattell advises the use of at least six tests to obtain each factor score. As the research situation in the Utah study did not permit the use of individual tests, six group tests were selected for each of the two factors. Tests selected for Factor UI-16 Competent Assertiveness were : G-6, G-20, G-27, G-42, G-44a, and G-45. For Factor UI-24, Anxiety to Achieve, tests G-26, G-29, G-31, G-38, G-41, and G-47 were selected. These tests employ a wide variety of approaches, but generally have no obvious relationship to the factor being measured and therefore are not fakable.

The Californina Psychological Inventory (CPI) contains 480 items and yields scores on 18 personality variables. The 19 scales are classified by the test author into four broad areas or classes. Cross-validational studies for each of the 18 scales of the CPI are summarized in the test manual (49).

Reliability coefficients on the 18 scales generally range around .70 and are to be made between groups rather than individuals.

Concurrent validities on the 18 scales are reported in a range from .26 to .83, with the majority yielding correlations in the .30 to .50 range with the concurrent criterion measures (49). The validity evidence for the CPI is substantial, although some of the validity coefficients are not as high as one might desire. When it is considered, however, that these coefficients are attenuated to a considerable degree by the reliability of both the scales and criterion measures and in many cases represent correlations with criterion measures that would not be expected to be highly reliable, the reported validities appear to be quite adequate for research purposes. Most of the concurrent validities are sufficiently high so that any marked influence of the grouping treatments upon these personality variables should be revealed in the analysis.

The Projective Measures

In the Utah study, the investigators wanted the spontaneous reactions of the children to pictures that portrayed situations relatively pertinent to the child and which were selected to reveal the variables to be measured in this research.

There were many tests and studies that were more or less pertinent to the needs of this study, but none of the tests completely met the specific testing needs, nor did any of the scoring methods serve as a model to follow. The final decision of the investigators was to select appropriate pictures from the variations of the TAT and other apperceptive measures available.

Cards from the various measures were evaluated by the investigators and members of the Department of Psychology at Utah State University. After considerable discussion and comparison, nine cards from the Thematic Apperception Test (TAT) (87) and the Michigan Picture Test (MPT) (83) were selected for use in the pilot study. Scoring procedures based on a five-point rating scale independently applied by two raters were developed in the pilot study. Procedures in the fifth-grade study called for individual administration of the measure with the entire testing session tape recorded. Typed transcripts of the recorded protocols were scored independently by the two raters. During the fifth-grade study, interrater reliability coefficients were regularly calculated on the blocks of protocols which had been scored between scorer conferences. This served as an objective check of the scorers' adherence to scoring rationale and procedures, and prevented them from drifting apart in their procedures and operational definitions. The composite rater reliabilities for the 144 protocols scored in the fifth-grade study were : aggression, .95; depression, .86; and inferiority feelings, .90.[15]

The sixth grade study, although building to a large extent upon the experience gained in the fifth grade study, required several procedural changes to adapt to the group testing situation. One new stimulus picture was employed, and the total number was reduced to seven. The pictures were presented to groups of sixth grade pupils by means of an opaque projector in a semi-darkened room. Pupils were required to write seven-minute stories for each of the seven TAT-type pictures, instead of responding orally as in the fifth-grade study. Having pupils write stories to TAT pictures instead of recording their responses according to the standard procedure considerably shortened the length of the stories elicited, and offered the raters fewer clues upon which to base a score.

The scoring method developed for the fifth grade study was applied to the sixth grade sample of 338 TAT-type story protocols, with the same two investigators scoring for both studies. The resulting interrater reliability coefficients obtained for the sixth grade sample were : aggression,.91; depression, .70; and inferiority, .73. These are lower than obtained in the fifth grade study, probably because of the shorter protocols obtained from the group administered test.

Sample

Several personality tests and related measures were employed in the Utah study. The samples to whom the various measures were administered are shown in Table 11. It will be noted that most of these measures were administered during the junior high school years. The lack of suitable measures appropriate for administration at the elementary school level reduced the amount of data that could be collected at this level. On the other hand, little personality data were collected at the high school level because of the limited differences between the grouping treatments as applied at the high school level in Districts A and R. As a rule, all subjects in a given sample were administered the personality measures scheduled for use with that sample, excepting the cases who were not available for either the regular or make-up testing. The projective measures, however, because of the expense of administration and scoring, were administered only to subgroups within Sample IV.

RESULTS

California Test of Personality

Between-treatment comparisons on the Feeling of Belonging subscores of this measure showed District R subgroups to be consistently higher than comparable District A subgroups. These differences reached statistical significance in three of the six comparisons, one at each ability level. Comparisons between pupils of different ability levels revealed a consistent tendency for pupils of higher ability to obtain more favorable scores on this subscale in both districts. These data suggest that pupils generally experienced a greater feeling of belonging in District R at all ability levels than in District A and also seem to indicate that superior pupils in both districts experienced a greater feeling of belonging than pupils of lower ability level. In interpreting these findings, the reader should bear in mind that, as the California Test of Personality was administered to Sample VI only, the findings may be accepted with somewhat less confidence than would have been the case if consistent results had been obtained for several samples. Also, the absence of concurrent validity evidence and the few items employed to measure each of the subscale areas would suggest that these results be interpreted with considerable caution.

Comparisons between District A and District R pupils on the Withdrawing Tendencies subscale revealed only one statistically significant difference favoring average boys in District R. The non-significant differences on this subscale were all in the direction of higher scores for District R pupils, thus presenting a rather consistent picture.

Between-treatment comparisons on the Anti-social Tendencies subscale revealed four significant differences in the six comparisons made, all favoring the District R groups. District R boys had significantly better scores at all three ability levels, while girls were significantly different only at the average ability level.

Inventory of Factors GAMIN

Two of the personality factors from this inventory, Ascendancy and Inferiority, were hypothesized to be influenced by differences in the grouping treatments. Between-treatment differences on the Ascendancy score were generally small. Of the 18 comparisons made between pupils in District A and District R on this factor, only two reached statistical significance. Both of the significant differences were found in Sample IV, one favoring superior District R boys and the other favoring slow District R girls. No consistent trends were apparent among the mean scores of comparable District A and R groups. Most of the differences between comparable groups were very small with 11 favoring District R groups and seven favoring District A groups. The lack of any consistent pattern in the differences obtained on this score plus the fact that only two significant differences were obtained from 18 comparisons seems to suggest that no true generalizable differences in the Acendancy score occurred as a result of differences between the two grouping situations.

Between-treatment differences on the Inferiority factor of the GAMIN were generally small. It was hypothesized that superior District A pupils would receive less favorable scores on this variable than comparable District R pupils. Although a slight trend was evident in the hypothesized direction, only one of the six comparisons between scores of superior pupils on this factor reached statistical significance. It was also hypothesized that slow pupils in District A would receive less favorable scores on this variable than their District R counterparts. The results showed no consistent pattern of differences favoring either grouping treatment for slow pupils on the Inferiority factor. These data would lead to the conclusion that feelings of inferiority as measured by GAMIN Factor I are not consistently different in ability grouped and random grouped classrooms.

OA Battery

Analysis of data obtained on the Objective-Analytic Personality Test battery required considerable effort. Cattell (25) recommends that the various tests going into a factor score be weighted equally, thus making it necessary to convert the raw scores on each test into standard scores and then add the standard scores for the subtests used in order to obtain a factor score. This procedure was followed to obtain a factor score for Factor UI-16, Competent Assertiveness. The performance of various Sample IV subgroups on this factor score was then compared, using analysis of variance.

In general, comparisons between similar pupils in the two grouping treatments revealed very similar means on this factor score. Only one of the six between-treatment comparisons revealed a statistically significant difference, District R slow girls obtaining a higher mean score ($t = 3.90$) than District A slow girls.

In order to gain some insights into the specific measures, the six tests selected to measure Cattell's factor UI-24, Anxiety to Achieve, were analyzed individually instead of combining them into a single factor score. Between-treatment comparisons on these six tests for superior pupils revealed not a single statistically significant difference among the 12 comparisons made. Similarly, comparisons for slow pupils failed to reveal a significant difference. Only the comparisons between average boys in the two districts revealed a consistent pattern of significant difference. Of the six between-treatment comparisons made for this group, four reached statistical significance, all favoring the District R pupils. One significant difference on Test G-47 was found in the between-treatment comparison of average girls, again favoring the District R sample.

Interpretation of the results on the Cattell factors is difficult, as would be expected from the exploratory nature of the measures involved. The general lack of consistent between-treatment differences, however, suggests that the factors these tests measure are not being differentially affected by the ability grouping and random grouping treatments.

California Psychological Inventory (CPI)

Between-treatment differences on the Class I variables of the CPI were generally small and inconsistent. The reader will recall that the six personality variables included in this class are broadly defined by the test author as measures of poise, ascendancy, and self-assurance. Of the 24 comparisons between superior pupils in the two districts, three reached statistical significance, all favoring Sample VII superior boys in District A over comparable District R pupils. As no significant differences occurred between superior pupils in Sample VI, the Sample VII results must be considered tentative. Comparisons between slow pupils in the two districts revealed only two significant differences, one favoring District R pupils and one favoring District A pupils.

Comparisons between average pupils revealed eight of the 24 between-treatment differences to be statistically significant. All eight of these differences indicated higher scores for District R pupils. Five of these differences occurred in Sample VI and three in Sample VII, suggesting some consistency in the results. No significant differences on these variables were hypothesized for average pupils and the investigator has no explanation for the fairly consistent tendency for District R pupils of average ability to attain more favorable scores on the Class I CPI personality variables.

Class III personality variables on the CPI include measures of achievement potential and intellectual efficiency. Three scores are obtained in this area: Achievement via conformance (Ac), Achievement via independence (Ai), Intellectual efficiency (Ie). It was hypothesized that both superior and slow pupils would make higher scores on Achievement via conformance in District A. One of the four comparisons between superior pupils on this variable reached statistical significance in the hypothesized direction. No significant differences between slow pupils were found. It was hypothesized that District A superior pupils would receive higher scores on Intellectual efficiency. Although District A superior pupils obtained higher scores on all four of the comparisons on this variable, only one of the differences approached significance.

It was hypothesized that superior District R pupils would obtain more favorable scores on Achievement via independence. The results, however, showed that District A superior pupils made higher scores on this variable in all four of the comparisons, one of which reached statistical significance. The only four between-treatment results that were reasonably consistent for the Class III variable were between superior boys in Sample VII. District A pupils in this sample were significantly higher in Ac and Ai and also obtained higher scores, approaching significance on Ie.

Class IV of the CPI contains three scores classified by the test author as measures of intellectual and interest modes. It was hypothesized that District R pupils at all ability levels would develop higher scores on flexibility than comparable District A pupils. Two significant differences were found between superior pupils on Flexibility. As one of these significant differences favors each treatment, they tend to cancel each other out, thus leading to rejection of the working hypothesis. No significant between-treatment differences emerged on the flexibility score for average and slow pupils.

The Projective Measures

The fifth grade study involved comparing the scores made by 71 District A slow pupils and 73 District R slow pupils on aggression, depression and inferiority feelings. None of the between-treatment differences were statistically significant.

In the sixth grade study pupils at all three ability levels were compared on the same three variables. Superior and slow pupils in the random grouped situation were expected to obtain higher mean scores on aggression and depression than similar pupils in ability grouped classrooms, as an effect of frustrations imposed by the school situation often considered to be geared primarily to the needs of the average pupil. This hypothesis was based on the assumption, derived from Mathias' (81) study, that the wide spread of individual differences within the heterogeneously grouped classroom would frustrate both the bright and slow pupils and cause a greater amount of aggression and depression in both groups. The only finding in keeping with this hypothesis was that District R slow girls demonstrated significantly more depression than slow girls in District A. On aggression, the only significant difference obtained was in the direction opposite that hypothesized, District A superior boys having scored significantly higher on this variable than comparable District R boys. The fact that the relationship is inverse, and does not hold for the slow pupils, casts serious doubt upon Mathias' interpretation of his findings. Mathias' (81) and Luchins and Luchins (72) findings applied to both girls and boys as a group. However, in the present findings a sex differential is present that suggests that inferences concerning the effects of a grouping practice upon any particular group of school children should consider the sex of the pupils as well as their level of academic achievement.

Regarding inferiority feelings, it was hypothesized that ability grouped slow pupils would react to being classified as "dumb" by obtaining higher mean scores on this variable than similar heterogeneously grouped pupils. The findings of the sixth grade study revealed no significant differences. Actually, both boys and girls received somewhat higher inferiority scores in District R, although these differences were well short of statistical significance. Thus, the claim that ability grouped slow pupils would react to the stigma of "being dumb", with increased feelings of inferiority, as advanced by Luchins and Luchins (72) was refuted. Data from the Utah study suggest that the method of grouping probably is not a significant factor in the development of feelings of inferiority among elementary school children. It might be that inferiority is a very basic characteristic and not as easily influenced as other more "superficial" personality variables. On the other hand, low ability pupils, grouped as such, may experience an increase in ego strength and feelings of adequacy due to less peer competition and greater chances for success. Ability grouping may also orient the pupil's parents to a better understanding of his potential and reduce parental expectation. It is possible that the aforementioned factors may offset any undesirable effect that the "dumb" label may have upon the slow pupil.

A final possibility that applies whenever negative results are found, is that the measures currently available are not sufficiently sensitive to reveal differences that might have developed in the two treatments.

SUMMARY AND CONCLUSIONS

California Test of Personality (CTP)

This measure was administered to Sample IV only.

It was hypothesized that superior and slow pupils in ability grouped classes would obtain higher scores on the Feeling of Belonging subscale. Results showed District R subgroups to be consistently higher on this score at all ability levels. These data suggest that ability grouping does not lead to a greater feeling of belonging on the part of pupils at any ability level but instead provides a less favorable climate than random grouping.

A lack of significant between-treatment differences on the Withdrawing Tendencies subscale suggests that this characteristic was not influenced by the grouping treatments.

Ability grouped pupils received less favorable scores on Anti-Social Tendencies in all six between-treatment comparisons, with four differences reaching statistical significance. The results suggest that ability grouped pupils, particularly boys, report significantly more Anti-Social Tendencies than do comparable pupils in random grouped classrooms. These differences were present at all three ability levels.

Inventory of Factors GAMIN

Two personality factors from this inventory, Ascendancy and Inferiority, were measured for pupils in Samples VI, VII, and VIII during the second year of the Utah study. These data, based on three samples, may be accepted with more confidence than the CTP data discussed above.

Of the 18 between-treatment comparisons made on the Ascendancy factor, only two reached statistical significance. These results suggest that the two treatments had no differential effect upon this personality variable.

No consistent between-treatment differences emerged on the Inferiority factor of the GAMIN. These results support the conclusion that ability grouping is no more likely to develop inferiority feelings in pupils at any ability level than is random grouping. This is particularly noteworthy in the case of slow pupils because of the frequently stated opinion that ability grouping does develop inferiority feelings among such pupils. These data, based on three different samples at different grade levels, are more extensive than any previous evidence published concerning this question.

Objective-Analytic Personality Test Factors (OA)

Tests measuring two factors from this battery were administered to Sample IV and Sample VI. No consistent pattern of between-treatment differences emerged on either of these variables. We may conclude that Competent Assertiveness and Anxiety to Achieve as measured on Cattell's OA battery were not affected differentially by the ability grouping and random grouping treatments.

California Psychological Inventory (CPI)

The CPI was admininstered to pupils in Samples VI and VII. The 18 CPI variables are classified into four broad areas. Class I variables include measures of poise, ascendancy, and self-assurance. Between-treatment comparisons on the Class I variables showed no consistent superiority for either treatment for superior pupils and slow pupils. Comparisons between average pupils revealed eight of the 24 between-treatment differences to be statistically significant, all favoring District R pupils. From these data we may conclude that the two grouping treatments do not appear to differentially affect personality variables related to poise, ascendancy, and self-assurance, excepting at the average ability level where random grouped pupils showed some tendency toward more favorable scores.

Class III variables on the CPI include three scores aimed at measuring Achievement via conformance (Ac), Achievement via independence (Ai), and Intellectual efficiency (Ie). Between-treatment differences on these three variables for subjects in Samples VI and VII were generally non-significant. There was some tendency for superior District A boys to obtain more favorable scores than comparable District R boys on these variables with two of the six comparisons between these groups reaching statistical significance and the remaining four differences following this same trend. With the possible exception of superior boys, it is concluded that the grouping treatments had no influence upon the broad area of achievement potential and intellectual efficiency as measured on the CPI.

Class IV of the CPI is concerned with intellectual and interest modes and is made up of three subtests: Psychological-mindedness (Py), Flexibility (Fx), and Femininity (Fe). No consistent between-treatment differences emerged on any of these three scores leading us to the conclusion that ability grouping and random grouping have no differential effect upon the Class IV CPI variables.

Projective Data

The first phase of the projective research, referred to as the fifth grade study, measured aggression, depression, and inferiority feelings on an individually administered TAT type projective measure using a sample of 144 slow pupils taken from ability grouped and random grouped fifth grade classes. The second phase, referred to as the sixth grade study, employed a similar group administered projective measure, using a sample of 338 slow, average and superior pupils taken from ability grouped and random grouped sixth grade classrooms.

Between-treatment comparisons on aggression, depression and inferiority feelings in the fifth grade study showed no statistically significant differences. Between-treatment differences on these three personality variables obtained in the sixth grade study also failed to show consistent significant differences between slow pupils on these three personality variables.

From the projective data it may be concluded that ability grouping and random grouping do not differentially affect the scores of slow pupils on aggression, depression and inferiority feelings. The lack of significant differences in inferiority feelings support the data obtained on other samples using the inferiority subscore of the GAMIN. Taken together, these results present strong evidence that ability grouping as employed in the Utah study does not cause the development of inferiority feelings among slow pupils.

The projective data obtained in the sixth grade study at all ability levels failed to support Mathias' (81) contention that aggression and depression are higher in random grouped classes for superior and slow pupils. The few significant between-treatment differences found on aggression and depression were not supported by data obtained in the fifth grade study. Thus, the most logical conclusion would be that these variables are not influenced by the ability grouping or random grouping treatments employed in the Utah study.

CHAPTER 12

A SUMMARY: CONCLUSIONS

THE GENERAL GOAL of this research has been to study differences in the effects upon elementary, junior high school, and high school pupils of an ability grouping system that differentiated the curriculum principally by adjusting the rate of presentation of curricular materials, and a random grouping system that differentiated the curriculum principally through the use of enrichment. Two adjacent and closely comparable school districts in Utah provided the setting in which these differences were explored. One of these districts employed random grouping with enrichment, and the other had adopted a system of ability grouping with acceleration, coincident with the start of this research.

Over 2,500 pupils from the fourth , sixth , seventh , eighth , and ninth grade levels were selected in the two districts at the beginning of the study. During the second year of the study, this sample was increased to about 4,000 pupils. Research data were collected over a four-year period in order to appraise the long term effects of the two grouping treatments. Pupils who were first tested early in grade four were followed through grade seven. As all samples were similarly followed over the period of the study, data were collected at all grade levels from four through twelve. The district that employed ability grouping is referred to throughout this report as District A. The other district participating in this research, known as District R, employed random or heterogeneous grouping.

Three major types of differences were considered in analyzing most of the data. The first and most important comparisons were between comparable pupils in the two grouping treatments. These comparisons involved pupils of the same sex and ability level in the ability grouped and random grouped samples and are referred to as between-treatment differences. These differences provide the best evidence on the differential effects of ability grouping and random grouping upon the dependent variables. Comparisons between pupils of the same sex and in the same grouping treatment who differed in ability level are referred to as level or within-treatment differences. These comparisons tell us whether the grouping treatments had a differential effect upon pupils at different ability levels. The final classification used in the analysis was sex. Boys and girls who were in the same grouping treatment and at the same ability level were compared to provide knowledge about the sex-related differences on the dependent variables.

ACHIEVEMENT

Achievement data were collected for all samples and gains from year to year as well as gains over the entire four years were analyzed. This procedure provided a large body of data on the nature of long term achievement in the two grouping treatments. Also, because achievement data for two or more samples were available at most grade levels, extensive cross checking of the results was possible. During the first year of the study, the California Achievement Test was used, while in subsequent years, measures of mathematics, science, social studies and reading from the STEP battery were administered.

Elementary

Achievement in four subject areas as well as overall achievement were compared during the elementary school years for these samples. A total of 54 statistical comparisons based on analysis of variance and covariance were made between ability grouped and random grouped pupils. Of these 54 comparisons, 28 revealed statistically significant differences. Nineteen of these differences were favorable to District A pupils while nine favored District R pupils. Of the 19 differences favoring ability grouped pupils, 15 occurred during the first year of the study. If these first year differences had been due to a substantial superiority of ability grouping over random grouping, we would have expected the differences to grow larger each year as the cumulative effects of the better system widened the achievement gap. No such cumulative effects occurred and in fact, many of the achievement differences favoring District A disappeared by the time Sample IV pupils had completed the sixth grade. Thus, we may conclude that neither ability grouping with acceleration nor random grouping with enrichment is more effective for all ability levels of elementary-school pupils. When data for the different ability levels were considered separately, achievement advantages of the two grouping systems, though small, tended to favor ability grouping for superior pupils and random grouping for slow pupils. As was hypothesized, the achievement results for average pupils did not consistently favor either grouping treatment.

Junior High School

All five of the samples employed in this study were in junior high school at some time during the four years of the project. When we combine the achievement analyses for these five samples, we find that a total of 60 statistical comparisons were made between comparable groups in the two grouping treatments; 33 in mathematics and 27 in science. Of the differences in mathematics achievement, five were significant in favor of District A subgroups, five in favor of District R, and the remaining 23 were nonsignificant. Of the science comparisons, five differences were significant favoring District A, one favoring District R and the remaining 21 were nonsignificant.

A study of the results at each of the three ability levels indicates that there was some tendency for ability grouping to lead to greater mathematics achievement among superior pupils and greater science achievement among average pupils. Among slow pupils, the differences between the two grouping treatments were small, but tend to favor District R in both mathematics and science.

OVERACHIEVERS AND UNDERACHIEVERS

In this phase of the Utah study, samples of fourth and sixth grade pupils from random grouped and ability grouped classrooms were classified as overachievers, underachievers and normal achievers based on the relationship between their achievement, chronological age, and mental age. The proportions of overachievers, normal achievers and underachievers among pupils of different ability levels in the two grouping systems were then compared.

There was a significant tendency for girls more frequently to be overachievers and less frequently to be underachievers than boys. This trend was present at both grade levels in both treatments.

Overall comparisons between District A and R indicated that there was a consistent trend for ability grouped pupils to be more often classified as overachievers and less often as underachievers. In comparing pupils at each of the three ability levels, we find that differences between superior pupils in the two districts were statistically significant at both fourth and sixth grades, with District A having more overachievers and a smaller percentage of underachievers than District R. No significant differences were found between the proportions of overachievers, underachievers, and normal achievers among pupils of average or low ability in the two districts.

STUDY METHODS AND ATTITUDES

Considerable evidence concerning study habits was collected over the four years of the Utah study. Pupils in Samples IV, VI, VII, and IX were administered study habits measured at grade levels ranging from seven to 12. Two measures were used, the California Study Methods Survey (CSMS) which yields three subtest scores, and the Survey of Study Habits and Attitudes (SSHA) which yield a total score only.

The Sample IV results on the CSMS showed all of the total score differences and seven of the nine subtest differences between comparable groups in Districts A and R to be statistically significant. Differences favored District R and were generally largest between pupils of superior ability. These results lead to the conclusions that pupils in a random grouping situation consistently develop better study methods during the elementary school years than pupils in an ability grouping situation.

Sample VI average pupils in District R received consistently higher scores on the CSMS than comparable District A pupils. No significant between-treatment differences were present for slow or superior pupils. Sample VIII average pupils in District R received higher SSHA scores. No significant between-treatment differences emerged from the Sample IX data. It may be concluded that if any advantage accrues from heterogeneous grouping at the secondary level, it is for average pupils, who made significantly greater gains in study habits in Samples VI and VIII.

SOCIOMETRIC CHOICE

A near-sociometric measure of the usual partial-rank-order type was developed for use in this study. Three criteria were employed, each with five positive choices. Each subject was asked to indicate by placing a check mark beside names on a roster of his classmates, those whom he preferred under each criterion. The three criteria were: 1) the five children in the class who are your best friends; 2) the five children with whom you prefer to study; and 3) the five children whom you would most like to have with you if you were transferred to another classroom. Each pupil's sociometric status was obtained by summing the choices he received on the three criteria. After obtaining a sociometric choice score for each pupil, the pupils were classified as stars, regulars, neglectees, or isolates.

Sample VI pupils were administered the sociometric measure only once – in the sixth grade – and therefore, this sample was used only in the study of sociometric status patterns. Sample IV pupils were measured in the fourth, fifth, and sixth grades, and data from this sample were used to study sociometric status patterns, mobility of sociometric status, and relationships between sociometric status and other variables.

The sociometric data collected seem to provide considerable evidence relating to sociometric choice in ability grouped and random grouped classes during the intermediate grades. Analysis revealed that the overall proportions of stars, regulars, and neglectees in District A and District R were similar, thus leading to the conclusion that ability grouping did not result in a permanent leadership vacuum in groups of average and slow pupils.

Superior students were found to lose some sociometric status when placed in ability grouped classrooms. This loss was particularly evident in the star classification but was not accompanied by any increase in the neglectee-isolate classification. Data on mobility of sociometric status suggest that superior pupils regain some of this status loss after they have made an adjustment to the ability grouped situation. Average and slow pupils appear to have a far better chance of gaining social recognition in ability grouped classrooms than do comparable pupils in random grouped classrooms. For the slow pupil, ability grouping not only appears to increase the pupil's chances of being classified as a star, but also reduces his chances of being classified as a neglectee-isolate. Sociometric status mobility patterns between District A and R demonstrate a greater number of status changes in ability grouped classrooms. The trend of these changes is generally toward higher sociometric status, further supporting the conclusion that organization of pupils into ability grouped classes brings about marked changes in sociometric status patterns during the first one or two years and generally leads to permanently higher status patterns for average and slow groups.

In random grouped classrooms, the pupil's ability level appears to be an important factor in determining his sociometric status. Sociometric status patterns differ considerably for pupils of different ability levels in random grouped classrooms, with pupils of higher ability consistently obtaining more favorable mean scores. The presence of significant correlations obtained when sociometric status was related to ability and achievement for District R pupils further demonstrates the importance of ability to status in the random grouped class. On the other hand, ability does not appear to be an important factor in determining sociometric status in ability grouped classrooms. In the Utah study, sociometric status patterns for pupils of different ability levels in ability grouped classrooms were closely comparable, showing no evidence of real differences. The absence of correlations between ability and status suggest that these factors are far less important to friendship choices in ability grouped classrooms.

We were also interested in learning what happened to a pupil's social status in the two grouping systems when his ability level changed. In District A such a change meant a change in classroom assignment, while in District R the pupil would be unaware of the change. First let us consider pupils who received a lower ability level classification in a succeeding year. District A pupils in this classification experienced significantly more upward mobility than was experienced by the comparable District R pupils. Thus, the downward movement seemed to increase the ability grouped pupil's relative social status as he was recognized as having come from a higher ability group and was near the top of his new group whereas he had been near the bottom of his previous group. These data would suggest that any damage to the ability grouped pupil's self-concept that might occur because of his reclassification to a lower ability section might be compensated to some extent by gains in social status he is likely to experience in his new class.

Pupils in District A and District R who moved upward in ability level classification were also studied. Two of the three chi-square comparisons made between these pupils yielded statistically significant differences. In each instance, pupils in random grouped classrooms showed more favorable sociometric status changes than comparable pupils in District A classrooms. These findings suggest that although ability grouping might initially provide greater opportunities for social leadership for those of average and low ability, it also might make more difficult the task of maintaining or gaining social recognition if the pupil makes achievement gains sufficient to raise his ability level classification. Loss of social status that is likely to accompany his reassignment to a new classroom might outweigh any gain in status he may experience from being identified as a member of the brighter group.

The lack of any consistent significant differences between boys and girls suggests that the effects of ability grouping and random grouping on sociometric variables are about the same for the two sexes.

Low but significant correlations were found to exist between sociometric status, pupil attitude, and self-concept variables. These correlations, though low, demonstrate the presence of relationships between the pupil's social acceptance and his perception of himself, his school environment, and his peers.

PUPIL ATTITUDES

Pupils in Sample IV were administered the USU School Inventory during their sixth grade year. This measure contains three scales: Attitude Toward Peers, Attitude Toward the Teacher, and Attitude Toward School.

Results on the Attitude Toward Peers scale revealed no significant between-treatment differences. Comparison on the Attitude Toward the Teacher scale showed that superior girls and boys and slow boys in District A received significantly more favorable scores than comparable District A pupils. No significant differences were found between average pupils in the two districts. Thus, ability grouping appears to be associated with more favorable attitudes toward the teacher among both superior and slow pupils.

Comparisons between the two treatments on the Attitude Toward School scale showed one significant difference favoring slow boys in District A. Within both treatments, there were significant differences between pupils of different ability levels all indicative of more favorable attitudes for pupils of higher ability. Three of the six sex differences of the Attitude Toward School scale were statistically significant. These differences, all of which favored girls, were largest at the superior level.

PUPIL PROBLEMS

The instruments employed to study pupil problems were the SRA Youth Inventory and the SRA Junior Inventory. These measures are essentially check lists each of which classifies the problems listed into a number of areas. A study of the problems listed on the two inventories revealed several areas in which one of the grouping treatments employed in the Utah study might be expected to differ from the other.

The SRA Junior Inventory, form S was administered to Sample IV pupils in District A and R during their fifth grade year. The SRA Youth Inventory, form A was administered to Sample VI pupils during their seventh grade year, Sample VII pupils during their eighth grade year, and Sample VIII pupils during their ninth grade year. As the SRA Youth Inventory was administered to three different samples at the junior high school level, the results on this measure provide fairly reliable evidence.

In reviewing the between-treatment differences on the problem areas of the Youth Inventory, perhaps the most noteworthy trend is the preponderance of significant differences indicating larger mean problem scores for the District R groups. A total of 32 between-treatment differences reached or approached statistical significance for the junior high-school samples. These differences were found in all samples. Of these 32 differences, only one indicated more favorable problem scores for District R pupils while the other 31 indicated more favorable problem scores for District A.

The frequency of significant differences between the districts was also found to be related to ability level. Of the 32 aforementioned differences on the SRA Youth Inventory, 15 were found between superior pupils in the two districts, and in 14 of these, the District R pupils obtained higher (less favorable) scores. Eleven of the 32 differences were between average pupils in the two districts; all indicating higher problem scores for District R groups. These results, reflecting as they do a consistently high number of problems for superior and average District R groups, would suggest the presence of true treatment differences at these ability levels. This trend, although present for slow pupils, was far less pronounced, and less general from sample to sample. These data support the overall conclusion that ability grouped junior high school pupils in average or superior groups report fewer problems than do comparable pupils in random grouped classes.

SELF-CONCEPT

The principal measure of self-concept employed in this research was the Index of Adjustment and Values, developed by Robert E. Bills (11). Pupils in all five samples were administered the IAV. Samples IV and VI were each tested three times over a three year period, while Samples VII, VIII and IX were each tested twice over a two year period. In the IAV the pupil answers three questions about himself: he rates himself with respect to each of the traits listed (concept-of-self), tells how he feels about being this sort of a person (acceptance-of-self), and how he would like to be with respect to each trait (ideal-self). In addition to yielding these three scores, a fourth score, Discrepancy, was obtained by subtracting the Concept of Self scores from the Ideal Self score.

For each of the four IAV variables, 66 between-treatment comparisons were made, 22 at each ability level.

Of the 66 comparisons on the Concept of Self score, 23 reached or approached statistical significance. All but two of these 23 differences were in a direction favoring District R pupils. Eight were between superior pupils in the two districts, ten between average pupils, and five between slow pupils. Fifteen of these 23 differences were between girls' groups while only seven were between boys' groups. The extent and consistency of differences in the Concept of Self data seem to justify the conclusion that at all ability levels, pupils in random grouped classrooms have more favorable concepts of self than comparable pupils in ability grouped classrooms. With respect to concept of self, the two grouping treatments had a somewhat greater effect upon girls than boys.

Of the 66 between-treatment comparisons of the IAV Acceptance of Self mean scores, 15 were statistically significant and seven others approached significance. Six of the significant differences were between superior pupils, six between average pupils, and three between slow pupils in the two districts. All 15 of these significant between-treatment differences favored District R as did four of the seven differences that approached significance. Based on these data, it must be concluded that random grouping is consistently related to higher self acceptance for pupils at all ability levels and over most of the grade levels covered in this project.

The overall summary of between-treatment comparisons on the IAV Discrepancy Score shows 14 of the 66 differences to be statistically significant. Six were between superior pupils in the two districts. Six of these differences were between average pupils. All but one of the 14 significant differences showed higher (less favorable) Discrepancy Score means for District A subgroups.

It may be concluded that the two grouping treatments have a differential effect upon the pupils' *Discrepancy Scores* with ability grouping generally associated with less favorable scores. This pattern is most pronounced in the differences between the mean scores of superior pupils in the two treatments, closely followed by average pupils. The differences between slow pupils are substantially smaller.

Of the 66 between-treatment comparisons of mean *Ideal Self* scores on the *Index of Adjustment and Values*, eight reached statistical significance. Seven of these eight differences favored District R pupils, with two occurring between superior groups, four between average groups and two between slow groups. The relatively few significant results suggests that the grouping system has little influence upon the *Ideal Self* score.

Sample IV pupils in District A who changed ability levels between the fifth and sixth grade administrations of the *IAV* were compared to determine whether changes in ability classification would be accompanied by changes in self-concept scores. The results indicated that remaining in the same ability classification in an ability grouping system appears to be less threatening to the pupil's self-concept than reclassification to either a higher or lower section. Placement of a pupil in a lower ability classification appears to bring about lower scores in *Concept of Self*, *Acceptance of Self*, and *Ideal Self*. Pupils destined to be reassigned to a lower ability classification were found to have lower scores in *Concept of Self* and *Acceptance of Self* prior to reassignment. This lower initial score may indicate that misassignment of a pupil to a higher group than his ability would justify brings about a lowering of his *Concept of Self* and *Acceptance of Self* because of his inability to perform satisfactorily in his assigned group.

PERSONALITY

Variables from four personality inventories were employed: The *California Test of Personality* (CTP); *Inventory of Factors GAMIN* (GAMIN) *Objective-Analytic Personality Tests* (OA); and the *California Psychological Inventory* (CPI). A projective measure employing cards from the *Thematic Apperception Test* (TAT) and the *Michigan Picture Test* (MPT) was also used to measure three variables that seemed particularly important.

California Test of Personality (CTP)

This measure was administered to Sample IV only. It was hypothesized that superior and slow pupils in ability grouped classes would obtain higher scores on the *Feeling of Belonging* subscales. Results showed District R subgroups to be consistently higher on this score at all ability levels. These data suggest that ability grouping does not lead to a greater feeling of belonging on the part of pupils at any ability level, but instead provides a less favorable climate than random grouping.

Ability grouped pupils also received less favorable scores on *Anti-social Tendencies* in all six between-treatment comparisons, with four differences reaching statistical significance.

Inventory of Factors GAMIN

Two personality factors from this inventory, *Ascendancy and Inferiority*, were measured for pupils in Samples VI, VII, and VIII during the second year of the Utah study. Of the 18 between-treatment comparisons made on the *Ascendancy* factor, only two reached statistical significance. These results suggest that the two treatments had no differential effect upon this personality variable. No consistent between-treatment differences emerged on the *Inferiority* factor of the *GAMIN*. Our findings support the conclusion that ability grouping is no more likely to develop inferiority feelings in pupils at any ability level than is random grouping.

Objective-Analytic Personality Test Factors (OA)

Tests measuring two factors from this battery were administered to Sample IV and Sample VI. No consistent pattern of between-treatment differences emerged for either of these variables. We may conclude that neither *Competent Assertiveness* nor *Anxiety to Achieve* as measured on Cattell's *OA* battery were affected differentially by the two grouping treatments.

California Psychological Inventory (CPI)

The *CPI* was administered to pupils in Samples VI and VII. The 18 *CPI* variables are classified by the test author into four broad areas. Class I variables include measures of poise, ascendancy, and self-assurance. Between-treatment comparisons on the Class I variables showed no consistent superiority for either treatment for superior pupils and slow pupils. Average pupils received more favorable scores in random grouped classes.

Class III variables on the *CPI* include three scores aimed at measuring *Achievement via conformance* (Ac), *Achievement via independence* (Ai) and *Intellectual efficiency* (Ie). Between-treatment differences on these three variables for subjects in Samples VI and VII were generally nonsignificant although there was some tendency for superior pupils in District A to obtain more favorable scores.

Class IV of the *CPI* is concerned with intellectual and interest modes and is made up of three subtests: *Psychological-mindedness* (Py), *Flexibility* (Fx), and *Femininity* (Fe). No consistent between-treatment differences emerged on any of these three scores.

Projective Data

The first phase of the projective research, referred to as the fifth grade study, measured aggression, depression, and inferiority feelings on an individually administered TAT type projective measure using a sample of 144 slow pupils taken from ability grouped and random grouped fifth-grade classes. The second phase, referred to as the sixth grade study, employed a similar group administered projective measure, using a sample of 338 slow, average, and superior pupils taken from ability grouped and random grouped sixth grade classrooms.

Between-treatment comparisons on aggression, depression, and inferiority feelings in the fifth grade study showed no statistically significant differences. Between-treatment differences on these three personality variables obtained in the sixth grade study also failed to show consistent significant differences.

The lack of significant differences in inferiority feeling support the data obtained on other samples using the inferiority subscore on the GAMIN. Taken together, these results present considerable evidence that ability grouping as employed in the Utah study does not cause the development of inferiority feelings among slow pupils.

OVERALL APPRAISAL OF THE GROUPING TREATMENTS

Let us now attempt to weigh the overall effects of the two grouping treatments on superior, average, and slow pupils at the elementary and junior high level. Table 20 provides a brief summary of these effects. Because of the limitations of the high school data collected in the Utah study, such an overall appraisal does not appear appropriate at that level. Any such appraisal of the two grouping treatments depends to a considerable degree upon the value system of the person making the appraisal. In other words, different persons would assign different amounts of importance to the various types of behavior and experience that were appraised by the measures in the Utah study. The following evaluation, therefore, reflects the value system of the investigator. Educators having different orientations may well draw different overall conclusions from the findings of this research.

Elementary

Superior Pupils: The superior pupil generally showed greater achievement gains in ability grouped classes. The differences in the Utah study were not large except for the first year, but data for the four years reflected a significant overall advantage for the ability grouping system. The data on over-achievers supported this finding, but as this phase of the research was carried out for only one year, the results must be considered tentative.

The study methods data for superior pupils showed a consistent and fairly large advantage for the random grouping treatment. Thus, although the ability grouped bright pupil may achieve more, he may develop less adequate study methods and be less capable in an independent study situation than his counterpart in the heterogeneous classroom.

When we consider the non-cognitive data for superior elementary school pupils, perhaps the most noteworthy finding is the loss in sociometric status and self-concept found for these pupils in the ability grouped situation. Some investigators have suggested that bright pupils have a somewhat inflated opinion of their own capabilities. If this is the case, perhaps the lower scores obtained on these measures by superior pupils in the ability grouping situation reflect a more healthy attitude than is found in the random grouped classroom. In elementary school classrooms, the teacher exerts considerable influence upon the peer group situation, and it may be that the teacher's tendency to place a very high value on achievement, results in high achieving pupils receiving a disproportionate amount of recognition. It will be recalled that the sociometric choice data, although indicating a reduction in the number of superior pupils of star status in the ability grouped classroom did not show an increase in the number of neglectee-isolates for this group. In weighing the total treatment effects for superior pupils during the intermediate grades in the elementary school, the investigator concludes that the ability grouping treatment is slightly more desirable.

Average Pupils: In reviewing the results of the Utah study for average pupils during the fourth, fifth, and sixth grades of the elementary school, we find a complex pattern of advantages and disadvantages associated with the two grouping treatments. In terms of achievement, it seems there is nothing to choose between the two treatments for average pupils. The average pupils, however, were found to have better study methods in the random grouped treatment.

With regard to non-cognitive variables, we find that average pupils showed more favorable personality characteristics, higher self-concept scores, and fewer pupil problems in the random grouped situation. The only aspect of their experience that strongly favored the ability grouping situation was the improved sociometric status they gained in this treatment. Although this improvement in status is not to be taken lightly, it seems that the preponderance of other variables favoring the random grouped situation must lead to the conclusion that the needs of most average pupils were better met in the heterogeneous classroom.

Slow Pupils: In terms of achievement related variables, the slow pupil generally showed better performance in the heterogeneous classroom. His

TABLE 20

SUMMARY OF COMPARISONS BETWEEN THE TWO GROUPING TREATMENTS

		Elementary			Junior High			
Variable	Sample	Sup.	Aver.	Slow	Sup.	Aver.	Slow	Remarks
Achievement	All	A*	N	R	A	A	R	Achievement differences over the four years were not large.
Overachievers and Underachievers	IV, VI	A	N	N				Based on first year's data only.
Study Methods and Attitudes	IV, VI VIII, IX	R	R	R	N	A	N	Elementary level differences largest for superior pupils.
Sociometric Choice	IV, VI	R	A	A				In ability grouped classrooms, gains of average and slow pupils are large.
Pupil Attitudes	IV							
Toward Peers		N	N	N				Slow boys had notably better attitudes in ability grouped classes.
Toward Teacher		A	N	A				
Toward School		N	N	A				
Pupil Problems	IV, VI VII, VIII	N	R	N	A	A	A	At junior high level, ability grouping was consistently related to fewer problems in all replications.
Self Concept Scores	All							
Concept of Self		R	R	R	R	R	R	Ability grouping was generally associated with less favorable self-concept scores at all levels and for all samples.
Acceptance of Self		R	R	R	R	R	R	
Ideal Self		N	R	N	N	N	N	
Discrepancy		R	R	R	R	R	R	
Personality								
1. Feeling of Belonging	IV	R	R	R				Personality measures at elementary level of doubtful validity.
2. Anti-Social Tendencies	IV	R	R	R				
3. Ascendancy	VI, VII VIII				N	N	N	The few significant differences on 3, 4, and 5 generally favored District R, but were not consistent between replications.
4. Inferiority	VI, VII VIII				N	N	N	
5. Competent Assertiveness	IV	N	N	R				
6. Anxiety to Achieve	VI				N	R	N	
7. CPI Class 1	VI, VII				A	R	N	Differences between average were more consistent than those between superior pupils.
8. CPI Class 3	VI, VII				A	N	R	
9. Flexibility	VI, VII				N	N	N	
10. Projective test								
Aggression	IV	N	N	N				
Depression	IV	N	N	N				
Inferiority	IV	N	N	N				

* A indicates results favoring ability grouping.
R indicates results favoring random grouping.
N indicates results favoring neither consistently.

achievement gains were rather consistently higher in the random grouped situation as were his study habits scores.

Review of the non-cognitive variables for slow pupils does not present a clear cut advantage for either grouping treatment. The most striking result of ability grouping for slow pupils was the tremendous gains they made in sociometric status. The slow pupil's attitudes toward both the school and teacher were also somewhat more favorable in the ability grouping situation. On the other hand, the self-concept data as well as the rather limited personality data available at this level tended to favor the random grouping. The reader may recall, however, that the differences between the two treatments in self-concept scores were much smaller for slow pupils than for superior or average pupils. Thus, it is the conclusion of the investigator that the large gains in sociometric choice are the most significant single effect of the grouping treatments on slow pupils. Therefore, it is concluded that the ability grouping system provides a more favorable environment for the slow pupil than does the random grouping treatment.

Junior High School

In reviewing Table 20, the reader will note that data on overachievers and underachievers, sociometric choice, and pupil attitudes were not collected at the junior high school level. Somewhat more personality data, however, were collected at this level, and these data can be accepted with somewhat more confidence than the personality data obtained during the elementary school years.

Superior Pupils: Ability grouping led to significantly greater achievement gains for superior pupils at the junior high school level, although these differences were not large. In reviewing the non-cognitive data, superior pupils in the ability grouped classes reported fewer pupil problems and obtained more favorable scores on the CPI measures of poise, ascendency and self-assurance (Class I), as well as on measures of achievement potential and intellectual efficiency (Class III). These differences on the CPI, however, were found in fewer than one-half of the replications, and may be regarded as tentative. When all data are considered, it appears to the investigator that ability grouping is slightly more advantageous for superior pupils at the junior high-school level.

Average Pupils: The pattern is somewhat the same for average groups, with ability grouped pupils making greater achievement and more favorable study methods scores. On the non-cognitive variables, average pupils in ability grouped classrooms reported fewer problems, but obtained somewhat less favorable self-concept scores than similar groups in the random grouped classrooms. Random grouped average pupils also obtained generally more favorable scores on the CPI Class I variables as well as on the Anxiety to Achieve measures on the Cattell OA battery. Again there seems little to choose between the two grouping treatments although the differences in the cognitive variables seem to give ability grouping a slight advantage.

Slow Pupils: Slow pupils in random grouped classrooms at the junior high school level generally achieved more than their ability grouped counterparts. Random grouped slow pupils also received somewhat more favorable self-concept scores although they reported a greater number of problems than the ability grouped samples. The personality data for slow pupils was characterized by a complete lack of consistent differences favoring either treatment. The lack of significant differences on aggression, depression, and inferiority feelings found in the projective phase of the personality study leads us to question some of the dire consequences that have been predicted by critics of ability grouping. Although personality data at this age level are at best tentative, the inferiority feelings and other negative personality characteristics thought to be caused by ability grouping of slow pupils did not emerge in the Utah study. In considering the overall pattern of differences, however, the investigator concludes that the heterogeneous classroom provides a slightly better environment for slow pupils at the junior high school level.

FOOTNOTES

1. The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

2. Most of the analysis of within-treatment and sex differences has not been included in this monograph, but may be found in the final research report prepared for the U. S. Office of Education: An Evaluation of Ability Grouping, Cooperative Research Project 577, Logan Utah: Utah State University, 1964. All aspects of the study are covered in a great deal more detail in this final report and the reader who requires such information is urged to refer to this report. All major findings related to the relative merits of the grouping treatments, however, are contained in this monograph.

3. A set of statistical hypotheses in the null from were developed as part of the original research plan.

4. For slow pupils, t = 2.09; for average, t = 2.44; for superior, t = 2.07.

5. A portion of this phase of the Utah Study is based on Dr. Nephi Manning's dissertation, which was completed under Dr. Borg's supervision.

6. Quoted from a personal communication from Dr. Miriam Goldberg, dated September 27, 1963, with the author's permission.

7. Quoted from a personal communication from Dr. Miriam Goldberg, dated September 27, 1963, with the author's permission.

8. Parts of this chapter are based on Dr. David Brinks' doctoral research carried out under Dr. Borg's supervision.

9. The three sociometric classifications used in this project were star, regular, and neglectee-isolate.

10. Portions of this section are based upon Mr. Val R. Christenson's Master's thesis which was carried out under Dr. Borg's supervision.

11. Data on these areas, as well as data on ability level differences, and sex differences may be found in the Office of Education, final report CRP577. These data were omitted from this monograph in order to conserve space.

12. Three other self-concept measures were also used in the Utah study: results may be found in the USOE Final Report, CRP577.

13. The Discrepancy Score was calculated by a different method in the Utah study than recommended by Bills.

14. The projective data are based upon the doctoral dissertations of Dr. Luna Brite and Dr. Clarence Swaner that were carried out under Dr. Borg's supervision.

15. Corrected using the Spearman-Brown prophecy formula.

REFERENCES

1. Adler, A. The Practice and Theory of Individual Psychology (New York: Harcourt Brace, 1924).
2. Armstrong, Marion E. A Comparison of the Interests and Social Adjustments of Underachievers and Normal Achievers at the Secondary School Level, PhD dissertation, University of Connecticut: University Microfilms, Ann Arbor, Michigan, 1955.
3. Baird, J. "Parallel Programs in Reading and Arithmetic," Detroit Educational Bulletin (1927), pp. 8-11.
4. Barbe, W. B. "Evaluation of Special Classes for Gifted Children," Exceptional Children, XXII (1955), pp. 60-62.
5. Barthelmess, H. M. and Boyer, P. A. "An Evaluation of Ability Grouping," Journal of Educational Research, XXVI (1932), pp. 284-94.
6. Bedoian, V. H. "Social Acceptability and Social Rejection of the Underage, at Age, and Overage Pupils in the Sixth Grade," Journal of Educational Research, XLVII (1954), pp. 513-20.
7. Bell, M. E. "A Comparative Study of Mentally Gifted Children Heterogeneously and Homogeneously Grouped," Dissertation Abstracts, XIX (1959), p. 2509.
8. Bettelheim, B. "Segregation: New Style," The School Review, LXVI (1958), pp. 251-272.
9. Billett, R. O. "A Controlled Experiment to Determine the Advantages of Homogeneous Grouping," Educational Research Bulletin, VIII (1928), pp. 133-40.
10. Billett, R. O. Provisions for Individual Differences, Marking and Promotion, Bulletin No. 17 (Washington, D.C.: U. S. Office of Education, 1932), Parts I, II, and III.
11. Bills, R. E., Vance, E. L., and McLean, O. S. "An Index of Adjustment and Values," Journal of Consulting Psychology, XV (1951), pp. 257-61.
12. Bonar, H. S. "Ability Grouping in the First Grade," Elementary School Journal, XXIX (1929), pp. 703-6.
13. Bonney, M. E. "A Study of the Relation of Intelligence, Family Size, and Sex Differences with Mutual Friendships in the Primary Grades," Child Development, XIII (1942), pp. 79-100.
14. Bonney, M. E. "The Relative Stability of Social, Intellectual, and Academic Status in Grades II to IV, and the Inter-relationships Between These Various Forms of Growth," Journal of Educational Psychology, XXXIV pp. 88-102.
15. Bonney, M. E. "Relationships Between Social Success, Family Size, Socioeconomic Home Backgrounds, and Intelligence Among School Children in Grades III to V," Sociometry, VII (1944), pp. 26-39.
16. Bower, E. M., Tashnovian, P. J., and Larson, C. A. "A Process for Early Identification of Emotionally Disturbed Children," Bulletin of California's State Department of Education, XXVII (1958), p. 18.
17. Bronfenbrenner, U. "The Measurement of Sociometric Status, Structure and Development," Sociometry Monographs, VI (1945).
18. Brookover, W. B., Velinsky, Ann, and Thomas, S. The Relationship of Self Image to Achievement in Junior High School Students, unpublished manuscript, Michigan State University, 1961.
19. Brown, W. F., and Holtzman, W. H. SSHA Manual, Revised (New York: Psychological Corporation, 1956).
20. Burdick, H. A. "Interrelationships of Three Measures of Motivation," Psychological Review, LXVIII (1961), p. 225.
21. Burtt, H. E., Chassel, L. M., and Hatch, E. M. "Efficiency of Instruction in Unselected and Selected Sections in Elementary Psychology," Journal of Educational Psychology, XIV (1923), pp. 154-61.
22. Buswell, M. M. "The Relationship Between the Social Structure of the Classroom and the Academic Success of the Pupils," Journal of Experimental Education, XXII (1953), pp. 37-52.
23. Carter, H. D. Manual, California Study Methods Survey (Los Angeles: California Test Bureau, 1958).
24. Carew, D. K. "A Comparison of Activities, Social Acceptance, and Scholastic Achievement of Men-Students," Personnel and Guidance Journal, XXXVI (1957), pp. 121-24.
25. Cattell, R. B. Handbook for the Objective-Analytic Personality Test Batteries (Urbana: Institute for Personality and Ability Testing, 1955).
26. Cavanaugh, M. H. "Grouping in Depth," Clearing House, XXXIV (1960), pp. 427-28.
27. Cook, R. R. "A Study of the Results of Homogeneous Grouping of Ability in High School Classes," Yearbook, National Society for the Study of Education, XXIII (1924), pp. 302-12.
28. Cunningham, R., et al. Understanding Group Behavior of Boys and Girls (New York: Columbia University, Bureau of Publications, 1951).

29. Daniels, J. C. "The Effects of Streaming in the Primary School," *British Journal of Educational Psychology*, XXXI (1961), pp. 69-78.

30. Dineen, M. A., and Gary, R. "Effects of Sociometric Seating in a Classroom Cleavage," *Elementary School Journal*, LVI (1955), pp. 358-62.

31. Drews, E. M. *The Effectiveness of Homogeneous and Heterogeneous Ability Grouping in Ninth Grade English Classes with Slow, Average, and Superior Students*, unpublished manuscript, Michigan State University, 1962.

32. Dvorak, A., and Rae, J. J. "Comparisons of Achievement of Superior Children in Segregated and Unsegregated First Grade Classes," *Elementary School Journal*, XXIX (1929), pp. 380-87.

33. Edmiston, R. W., and Benfer, J. C. "Relationship Between Group Achievement and Range of Abilities Within Groups," *Journal of Educational Research*, XLII (1949), pp. 547-48.

34. Ekstrom, Ruth B. *Experimental Studies of Homogeneous Grouping, A Review of the Literature* (Princeton, N. J.: Educational Testing Service, 1959).

35. Fliegler, L. A. *Psychology*, 3rd edition (New York: World Book Co., 1957).

36. Fowlkes, J. G. "Homogeneous or Heterogeneous Grouping—Which?" *Nation's Schools*, VII (1931), pp. 74-78.

37. French, J. W. "Evidence From School Records on the Effectiveness of Ability Grouping," *Journal of Educational Research*, LIV (1960), pp. 83-91.

38. Freud, S. *The Ego and the Id* (London: Hogarth Press, 1950).

39. Gallagher, J. J. "Peer Acceptance of Highly Gifted Children in Elementary School," *Elementary School Journal*, LVIII (1958), pp. 465-70.

40. Gibboney, R. A. "Socioeconomic Status and Achievement in Social Studies," *Elementary School Journal*, LIX (1959), pp. 340-46.

41. Ginzberg, E. *Values and Ideals of American Youth* (New York: Columbia University, 1961).

42. Goldberg, Miriam L. "Research on the Talented," *Teachers College Record*, LX (1958), pp. 150-63.

43. Goldberg, Miriam L. "Research on the Gifted," in Shertzer (Ed.,) *Working with Superior Talented Students* (Chicago: Science Research Associates, 1960).

44. Goldberg, Miriam L., et al. "A Three Year Experimental Program at DeWitt Clinton High School to Help Bright Underachievers," *High Points*, XLI (1959), p. 1.

45. Goldberg, Miriam L., et al. *The Effects of Ability Grouping*, unpublished manuscript, Teachers College, Columbia University, 1961.

46. Goldberg, Miriam L., et al. "Social Consequences of Special Education for the Talented," *The Education Digest*, XXIV (1958), pp. 11-14.

47. Goldworth, M. "The Effects of an Elementary School Fast-Learner Program on Children's Social Relationships," *Exceptional Children*, XXVI (1959), pp. 59-63.

48. Goodlad, J. I. "Pressures Can Be Blocks to Learning, Don't Push Me," *Bulletin, Association for Childhood Education International* (1960), pp. 23-28.

49. Gough, H. G. *Manual for the California Psychological Inventory* (Palo Alto, Calif.: Consulting Psychologists Press, 1957).

50. Gowan, May Seagoe. "Why Homogeneous Grouping?" *California Journal of Secondary Education*, XXX (1955), pp. 22-28.

51. Gray, H. A., and Hollingsworth, L. S. "The Achievement of Gifted Children Enrolled and Not Enrolled in Special Opportunity Classes," *Journal of Educational Research*, XXIV (1931), pp. 255-61.

52. Gronlund, N. E. *Sociometry in the Classroom* (New York: Harper and Brothers, 1959).

53. Gronlund, N. E., and Whitney, A. P. "Relation Between Pupil's Social Acceptability in the Classroom, in the School, and in the Neighborhood," *The School Review*, LXIV (1955), pp. 267-71.

54. Guilford, J. P., and Guilford, R. B. "Personality Factors D, R, T, and A," *Journal of Abnormal Social Psychology*, XXXIV (1939), pp. 21-36.

55. Guilford, J. P., and Martin, H. G. *The Guilford-Martin Inventory of Factors GAMIN, Manual of Directions and Norms* (Beverly Hills, Calif.: Sheridan Supply Co., 1948).

56. Hamalainen, A. E. "Method of Grouping Pupils Should Provide Normal Social Situations," *Nations Schools*, XLV (1950), pp. 34-35.

57. Hansen, C. F. *A Third Year Evaluation Report on the Four Track Curriculum*, unpublished manuscript, Board of Education, District of Columbia, April 15, 1959.

58. Hartill, R. W. *Homogeneous Grouping* (New York: Bureau of Publications, Teachers College, Columbia University, 1936).

59. Herr, W. "Junior High School Accelerates and Their Peers in Senior High School," *The School Review*, XLV (1937), pp. 186-95, 289-99.

60. Hinkelman, E. A. "Relationship of Intelligence to Elementary School Achievement," *Educational Administration and Supervision*,

XLI (1955), pp. 176-79.

61. Hollingshead, A. B., and Redlich, R. C. Social Class and Mental Illness (New York: John Wiley and Sons, 1958).
62. Horney, K. The Neurotic Personality of Our Times (New York: W. W. Norton, 1937).
63. Jackson, G. T. "Each According to His Ability," School Executive, LXII (1943), pp. 37-38.
64. Jensen, V. H. "Influence of Personality Traits on Academic Success," Personnel Guidance Journal, XXXVI (1958), pp. 497-500.
65. Jersild, A. R. In Search of the Self (New York: Bureau of Publications, Teachers College, Columbia University, 1952).
66. Jones, V. A., and McCall, W. A. "Application of Two Techniques in Evaluating Some Policies of Dealing with Bright Children," Teachers College Record, XXVII (1926), pp. 825-35.
67. Justman, J. "Academic Achievement of Intellectually Gifted Accelerants and Nonaccelerants in Junior High School," School Review, LXII (1954), pp. 142-50.
68. Kravaceus, W. C., and Wiles, E. "An Experiment in Grouping for Effective Learning," Elementary School Journal, III (1938), pp. 264-68.
69. Laughlin, F. The Peer Status of Sixth and Seventh Grade Children (New York: Bureau of Publications, Teachers College, Columbia University, 1954).
70. Lawson, D. E. "An Analysis of Historic and Philosophic Considerations for Homogeneous Grouping," Educational Administration and Supervision, XLIII (1957), pp. 257-70.
71. Lecky, P. Self-Consistency: A Theory of Personality (New York: Island Press, 1945).
72. Luchins, A. S., and Luchins, E. H. "Children's Attitudes Toward Homogeneous Grouping," Pedagogical Seminary and Journal of Genetic Psychology, LXXII (1948), pp. 3-9.
73. MacLean, M. S. "What and Why of Underachievement," The High School Journal, XLII (1958), pp. 69-72.
74. Mann, H. "How Real are Friendships of Gifted and Typical Children in a Program of Partial Segregation?" Exceptional Children, XXIII (1957), pp. 199-201.
75. Mann, M. "What Does Ability Grouping Do to the Self Concept?" Childhood Education, XXXVI (1960), pp. 357-60.
76. Marsh, D. E. "An Interesting Experience in Grouping," The California Journal of Secondary Education, XXX (1955), pp. 49-59.
77. Martin, W. H. The Results of Homogeneous Grouping in the Junior High School, unpublished PhD dissertation, Yale University, 1927.
78. Martinson, Ruth A. "The California Study of Programs for Gifted Pupils," Exceptional Children, XXXVI (1959), pp. 339-43.
79. Martinson, Ruth A. Study Project on Programs for Gifted Pupils, unpublished manuscript, California State Department of Education, Sacramento, 1960.
80. Maslow, A. H. Motivation and Personality (New York: Harper, 1954).
81. Mathias, R. Methodology and Results of Personality Appraisal with Rorschach, TAT and Figure Drawing, paper read at the American Psychological Association annual meeting, Cincinnati, September 1959.
82. McGaughy, J. R. "Homogeneous Grouping of Pupils," Childhood Education, VI (1930), pp. 291-96.
83. Michigan State Department of Mental Health. Michigan Picture Test (Chicago: Science Research Associates, 1953).
84. Mosier, C. I. "A Factor Analysis of Certain Neurotic Tendencies," Psychometrika, II (1937), pp. 263-87.
85. Moyer, E. L. "A Study of the Effects of Classification by Intelligence Tests," Yearbook, National Society for the Study of Education, XXIII (1924), Part I.
86. Muggenthaler, E. "The Sociological Status of Slow-Learning Pupils in an Elementary Grade," Prax, Kinder-Psycholgie Kinder-Psychiatrie, IV (1955), pp. 226-30.
87. Murray, H. A. Thematic Apperception Test Manual, and Thematic Apperception Test (Cambridge, Mass.: Harvard University, 1943).
88. Otto, H. J. Elementary School Organization and Administration (New York: Appleton-Century-Crofts, 1954).
89. Passow, A. H., Goldberg, Miriam L., and Link, F. R. "Enriched Mathematics for Gifted Junior High School Students," Educational Leadership, XVIII (1961), pp. 442, 448.
90. Provust, M. M. "Ability Grouping in Arithmetic," Elementary School Journal, LXX (1960), pp. 391-98.
91. Purdom, R. L. A Scientific Study to Determine the Value of Homogeneous Grouping Made on the Basis of Intelligence Tests, unpublished PhD dissertation, University of Michigan, 1924.
92. Rankin, P. "Evaluation of Ability Grouping in the Seventh Grade," Review of Educational Research, I (1931), pp. 223-25.
93. Rankin, P., Anderson, C. T., and Bergman, W. G. "Ability Grouping in the Detroit Individualization Experiment," Yearbook, National Society for the Study of Education, XXXV (1936), Part I.
94. Remmers, H. H., Shimberg, B., and Drucker, A. J. Technical Supplement for the SRA Youth Inventory, Form A (Chicago: Science

Research Associates, 1953).

95. Remmers, H. H., and Bauernfeind, R. H. Manual for the SRA Junior Inventory, Form S (Chicago: Science Research Associates, 1955).
96. Roberts, I. L. "Homogeneous Grouping: Experiment," Chicago Schools Journal, XXIX (1947), pp. 30-32.
97. Rochfort, G. B., Jr. "Evaluating Social-Personal Outcomes in Differentiated Instruction," Journal of Education, CXLII (1959), pp. 66-71.
98. Rogers, C. R. Client-Centered Therapy (Boston: Houghton-Mifflin, 1951).
99. Rosenfeld, H., and Zander, A. "Teachers' Influence on Student Aspirations," Journal of Educational Psychology, LII (1961), pp. 1-11.
100. Russell, D. H. "Inter-Class Grouping for Reading Instruction in the Intermediate Grades," Journal of Educational Research, XXXIX (1946), pp. 462-70.
101. Russell, D. W. "A Plea to Beam in the Underachiever," The High School Journal, XLII (1958), pp. 66-68.
102. Ryan, R. J., and Davie, J. S. "Social Acceptance, Academic Achievement, and Academic Aptitude Among High School Students," Journal of Educational Research, LII (1958), pp. 101-6.
103. Saville, A. "A Survey of Ability Grouping," Dissertation Abstracts, XXII (1961), pp. 22, 1473.
104. Sears, Pauline S. "Levels of Aspiration in Academically Successful and Unsuccessful Children," Journal of Abnormal and Social Psychology, XXXV (1940), pp. 489-526.
105. Simpson, R. E., and Martinson, Ruth A. Educational Programs for Gifted Pupils (Sacramento, Calif.: California State Department of Education, January 1961).
106. Snygg, D., and Combs, A. W. Individual Behavior: A New Frame of Reference for Psychology (New York: Harper, 1949).
107. Staker, A. M. "Changes in Social Status of Elementary-School Pupils," Educational Research Bulletin, XXVII (1948), pp. 157-59.
108. Stivers, E. H. "Motivation for College in High School Girls," The School Review, LXVII (1959), pp. 320-34.
109. Stone, D. R., and Ganung, G. R. "A Study of Scholastic Achievement Related to Personality as Measured by the Minnesota Multiphasic Personality Inventory," Journal of Educational Research, L (1956), pp. 155-56.
110. Symonds, P. M., and Jackson, C. E. Measurement of the Personality Adjustment of High School Pupils (New York: Teachers College, Columbia University, 1935).
111. Theissen, W. W. "The Relative Progress of Groups Sectioned on the Basis of Ability," Journal of Educational Research, V (1922), pp. 295-305.
112. Thorpe, J. G. "An Investigation into Some Correlates of Sociometric Status Within School Classes," Sociometry, XVIII (1955), pp. 49-61.
113. Tonsor, C. A. "Must We Segregate?" National Association of Secondary School Principals Bulletin, XXXVII (1953), pp. 75-77.
114. Ullman, C. A. "Identification of Maladjusted School Children," Public Health Monograph, No. 7 (1952).
115. Ullrich, O. A. An Experimental Study of the Effect of Learning on Sectioning College Classes on the Basis of Ability, unpublished PhD dissertation, University of Texas, 1926.
116. Van Wagenen, M. J. "Effects of Homogeneous Grouping Upon the Quality of Work of Superior Children," Educational Methods, VI (1927), pp. 240-47.
117. West, J., and Sievers, C. "Experiment in Cross-Grouping," Journal of Educational Research, LIV (1960), pp. 70-72.
118. Westmayer, P. "Just Stand Out of the Way!" School Science and Mathematics, LVII (1957), pp. 643-46.
119. Wilson, Robert C. The Gifted Children, Portland (Portland, Oregon: Portland Public Schools, 1959).
120. Winkler, J. A., and MacNutt, T. Research Relating to Children, Bulletin No. 11, U.S. Department of Health, Education, and Welfare (Washington, D.C.: U. S. Government Printing Office, 1960).
121. Woody, C. "Results Obtained from Ability Grouping," Bureau of Educational Reference and Research Bulletin, 1924.
122. Worlton, J. T. "The Why of Homogeneous Classification," Elementary School Journal, XXVII (1926), pp. 265-74.
123. Worlton, J. T. "The Effect of Homogeneous Classification of Scholastic Achievement of Bright Pupils," Elementary School Journal, XXVIII (1928), pp. 336-45.
124. Wrightstone, W. J., Justman, J., and Robbins, I. Evaluation in Modern Education (New York: American Book Co., 1956), pp. 199-214.
125. Wylie, Ruth. The Self Concept (Lincoln, Neb.: University of Nebraska Press, 1961).